UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

HYAGO SALLET

Simplified Literature Review on the Applicability of Process Mining to RPA

Work presented in partial fulfillment
of the requirements for the degree of
Bachelor in Computer Science

Advisor: Profa. Dra. Lucineia Heloisa Thom

Porto Alegre
May 2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
Reitor: Prof. Carlos André Bulhões Mendes
Vice-Reitora: Profª. Patrícia Helena Lucas Pranke
Pró-Reitora de Ensino: Profª. Cíntia Inês Boll
Diretora do Instituto de Informática: Profª. Carla Maria Dal Sasso Freitas
Coordenador do Curso de Ciência de Computação: Prof. Rodrigo Machado
Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

ABSTRACT

Business processes play an important role in any enterprise value chain and are involved in key activities such as the purchase of material, sales, and hiring of staff. Hence, medium-sized and large companies are inherently process-oriented. Managing business processes is yet, due to new regulations, technologies, and market changes, not a trivial task. In addition to that, the execution of business processes may be repetitive, tedious and time demanding. For this reason, there is a high motivation to automate such processes, which has been facilitated by the popularisation of Robotic Process Automation (RPA). RPA brings a cost-efficient solution for process automation along with a substantial challenge that is to decide what process to automate and how. Process Mining tools and techniques have been largely adopted to address challenges faced during RPA implementations. The goal of this work is to present the usage of Process Mining in RPA implementations through a simplified systematic literature review.

Keywords: Business Process. Process Mining. Robotic Process Automation. RPA.

RESUMO

Processos de negócio possuem um papel importante em qualquer cadeia de valores corporativa e estão envolvidos em atividades chave como compras de suprimentos, vendas e contratações de recursos humanos. Por esse motivo, empresas de médio e grande porte são inerentemente orientadas a processos. Devido à novas regulamentações, tecnologias e mudanças de mercado, a gestão de processos de negócio é ainda uma tarefa não trivial. Além disso, a execução de processos de negócio pode ser repetitiva, entendiante e demandar tempo. Por isso, existe uma alta motivação para automatizar processos de negócio, o que tem sido facilitado pela popularização da Automação de Processos Robóticos (Robotic Process Automation - RPA). RPA provê uma solução eficiente em custo para automação de processos e trás desafios no âmbito das escolhas de quais precessos automatizar e como. As ferramentas e metodologias de Mineração de Processos têm sido amplamente utilizadas para endereçar os desafios provenietes de implementações de RPA. O objetivo deste trabalho é apresentar as aplicações da Mineração de Processos em RPA, através de uma revisão sistemática simplificada da literatura.

Palavras-chave: Business process, Process Mining, Robotic Process Automation,RPA.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	*Artificial Intelligence*
BPM	*Business Process Management*
BPMN	*Business Process Model and Notation*
BPMS	*Business Process Management System*
CRM	*Customer Relationship Management*
ERP	*Enterprise Resource Planning*
IT	*Information Technology*
PAISs	*Process-Aware Information Systems*
PLM	*Product Lifecycle Management*
PO	*Purchase Order*
RPA	*Robotic Process Automation*
SCM	*Supply Chain Management*
SSLR	*Simplified Systematic Literature Review*
SW	*Software*
UI	*User Interface*
WFM	*Workflow Management*

CONTENTS

1 INTRODUCTION

A business process is a collection of linked tasks, which aims at delivering a service or product to a customer, or to accomplish an organizational goal (DUMAS et al., 2018). As a result, business processes has become key success criterion to large enterprise, which are constantly interested in improving and managing their existing business processes. For that reason, Business Process Management (BPM) topics stays in evidence both in the industry of enterprise software and in the academia (AALST et al., 2016). Furthermore, the recent popularisation of AI technologies and automation tools has added a catalyst to BPM field (PASCHEK; LUMINOSU, 2017). With the adoption of Process Mining (AALST, 2012), former complex BPM activities such as process discovery, process re-design and monitoring has turned into significantly less time consuming and tedious tasks. As a result, companies tend to become more process-aware and are able to focus efforts on standardizing and optimizing their respective business processes in a competitive speed. The advances in BPM, resulted from the application of AI and Process Mining, facilitates automation. By standardizing and optimising business processes, it is possible to take advantage from Robotic Processes Automation (RPA) (HOFMANN; SAMP; URBACH, 2020) to reduce the need of dedicating expensive human resources to the execution of repetitive and low-value tasks within a line of business. Besides the reduction of expensive resources allocated to the execution of repetitive tasks, RPA can support companies on reallocating people and developing them to participate in high-value tasks such as strategic decisions to their respective areas of expertise.

1.1 Motivation

BPM is subject of study for over two decades, and recently has gained visibility from different optics due to the increasing amount of Process Mining and RPA tools available. In the past decade, BPM was most utilized for modelling business processes with zero to non-technical integration to the process execution, or used for redesigning existing business processes models which do not reflect the real process being executed by computer systems (AALST, 2020). Despite of the recent use of RPA and Process Mining tools in the BPM field, the scientific research on those topics is yet novel (IVANCIC; VUGEC; VUKSIC, 2019). Currently RPA and Process Mining implementations are mostly based on its respective solution vendors best practices and not in the academic literature. The motivation of this work is to present the usage of Process Mining in RPA implementations and answer research questions, which might be helpful for future academic researches on RPA and Process Mining. This work can be used as a starting point for future researches looking for evidences of Process Mining applicability in RPA.

1.2 Goals

The goal of this work is to present the usage of Process Mining techniques applied to RPA through a Simplified Systematic Literature Review (SSLR). The SSLR aimed at selecting papers, which describes implementation stages of RPA and application of Process Mining tools and techniques on the respective RPA project phases.

1.3 Research Questions

Considering our research goal, we defined three research questions, according to the guidelines to Systematic Literature Review (KITCHENHAM et al., 2009) presents:

RQ1 – *What are the implementation phases of RPA?*

RQ2 – *In which implementation phases of RPA is Process Mining being applied?*

RQ3 – *What are the Process Mining tools and techniques being used in the implementation of RPA?*

1.4 Text Organization

This work is organized as follows. Chapter 2 presents the fundamentals for the conducted research. It consists of an overview of BPM, along with an overview of Process Mining and RPA. Chapter 2 also presents an overview of related work. Chapter 3 presents the SSLR protocol. It contains the research questions, the selection criteria and the research procedures. Chapter 4 brings the results from the data collected through the SSLR. This Chapter presents the results and subsequently a methodological analysis answering the proposed research questions. Finally, Chapter 5 summarises the findings, and presents future directions for this research.

2 FUNDAMENTALS

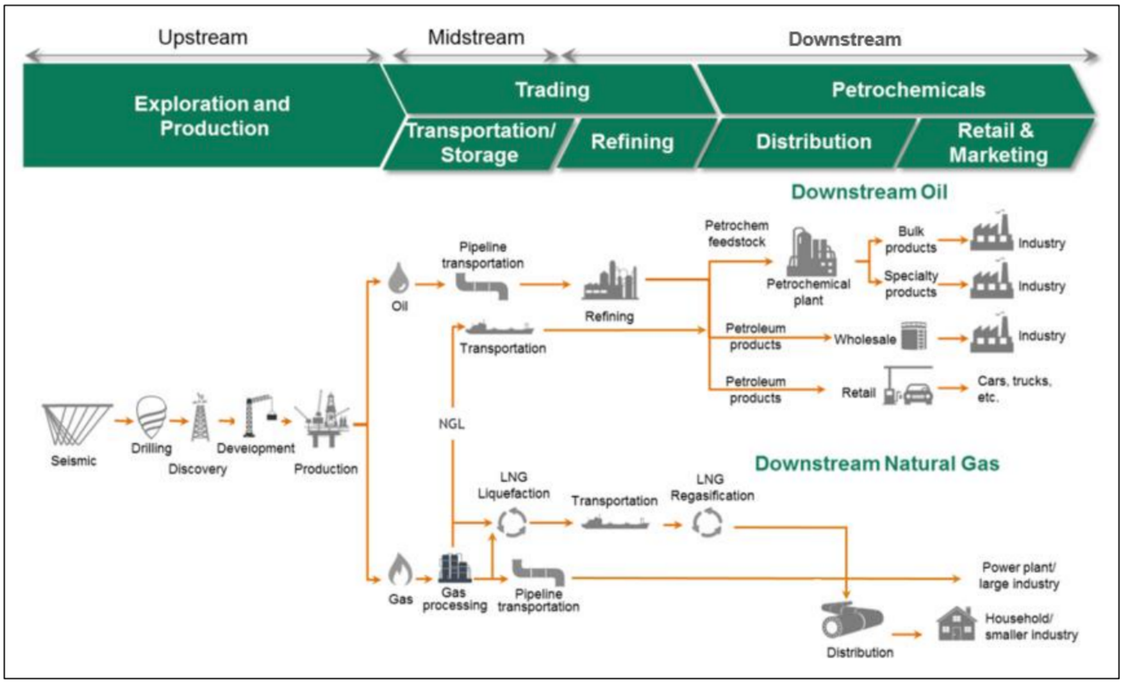
This chapter provides the background knowledge necessary for the development of this work. At first it presents an overview of BPM. Second, it describes Process Mining and its respective techniques (e.g. process discovery, process conformance and process enhancement). Third, it defines RPA and its current research challenges. In summary, this chapter mainly intends to connect the BPM field with Process Mining by presenting the process discovery phase of BPM, while it connects BPM with RPA by describing the concept of Process-Aware Information Systems (AALST, 2009).

2.1 Business Process Management

The presence of business processes is inevitable to every single organization or enterprise. Essentially, each product or service delivered by a certain line of business is ruled through a set of activities and tasks (DUMAS et al., 2018). As a real life example, we can use a well known end-to-end business processes such as the Oil & Gas industry, which involves activities such as Exploration and Production, Transportation/Storage, Refining, Distribution and, finally, Retail and Marketing. In the figure 2.1, it is illustrated the several activities and its respective tasks in the Oil & Gas value chain. The final product, Oil or Gas, arrives at the final customer after going through a series of business process steps. Business processes, such as the one represented by the Oil & Gas industry, are usually described with the use of Business Process Model and Notation (BPMN) (BULE et al., 2015).

A business process is a set of events, activities, and decisions which aims at adding value to a organization, and its customers. (DUMAS et al., 2018) states that BPM can be considered the art and science of overseeing how work is performed in an organization with the goal of ensuring consistent outcomes and to take advantage of improvement opportunities.

Figure 2.1: Figure illustrating the end-to-end Business Process of Oil & Gas Value Chain

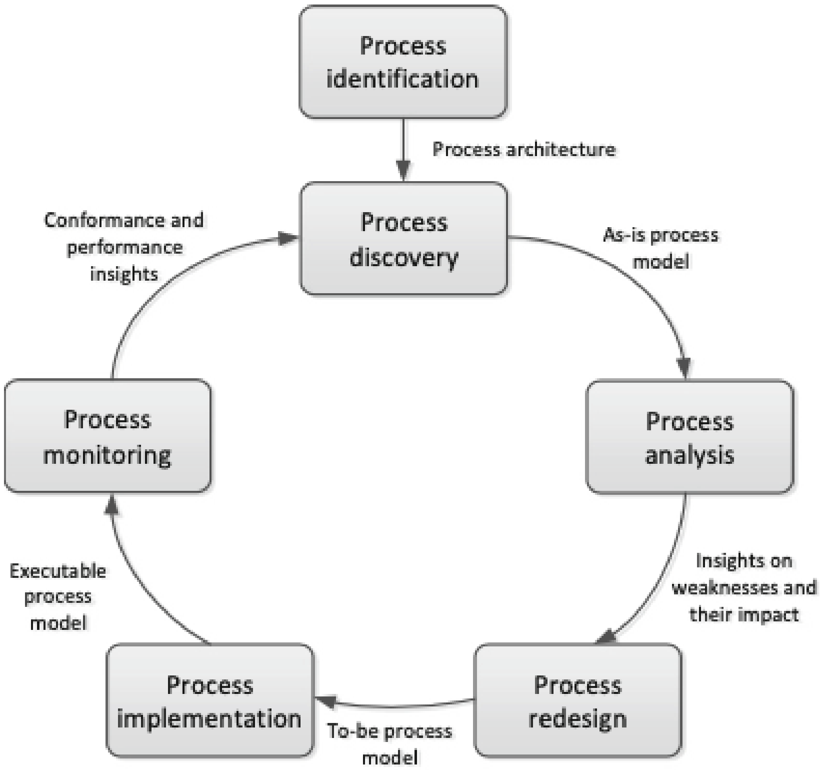


Source: (ÁLVAREZ et al., 2018)

2.1.1 Business Process Management Lifecycle

The BPM lifecycle consists of six continuous stages aiming at assisting BPM initiatives in organizations. Each BPM cycle represents a phase with well established goals of identifying, discovering, analysing, redesigning, implementing and monitoring business processes. Additionally, the BPM lifecycle counts with a set of methods and tools to identify and manage processes. These tools and methods are usually incorporated in Business Process Management System (BPMS) (KARAGIANNIS, 1995) for graphical representation of the resulting processes discovered and managed through the BPM lifecycle. Figure 2.2 illustrates the BPM lifecycle.

Figure 2.2: The BPM lifecycle



Source: (DUMAS et al., 2018)

Process Identification

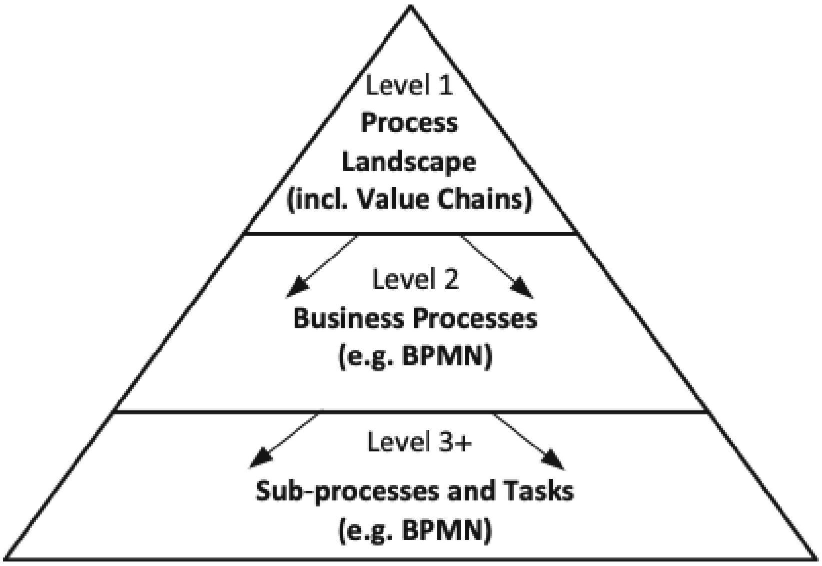
One of the main goals of BPM field is to improve process of an organization. In order to achieve that, it is necessary to start by recognizing the business processes, which should be prioritized for improvements. The systematically methodology and criteria for selecting specific processes for improvement is called *process identification*.

Process identification uses a systematic method for identifying business processes. It builds an organization's process architecture, which consists of three levels of abstraction. They are categories of processes, relationships between processes and the definition of the processes landscape (POLANČIČ et al., 2020).

The categories of processes defines the business processes which are part of the core business of an organization, support the business or are associated with management area. After this, it is establish relationships of sequence, decomposition, and specialization of these processes.

The result of the identification phase consists of a high-level description of processes, which are part of an organization's different areas, followed by a second level containing specific information and relationships between each of the processes identified in the initial abstract model. Finally, the third level presents the processes with additional details in a granularity of sub-processes and tasks.

Figure 2.3: Process architecture resulted from the Process Identification



Source: (DUMAS et al., 2018)

Process Discovery

Process discovery is an activity which focus on gathering information about existing process and organizing it in terms of an as-is process model, which represents the current state of a process being implemented by an organization (DUMAS et al., 2018). Differently from the Process Discovery in the context of Process Mining (presented in section 2.2.1 of this work), BPM process discovery mostly requires manual gathering and organizing of the business process information. Often resulting in a tedious and time-consuming activity. Hence, several organizations fail to have business processes well documented. In order to maximise the effectiveness of manual business process discovery, (DUMAS et al., 2018) proposes the following four tasks of process discovery:

1. Defining the setting: Elect a team in an organization who are responsible for working on the process.
2. Gathering information: Focus on building an understanding of the processes being discovered. A set of process discovery methods can be utilized for that purpose.

3. Conducting the modeling task: The modeling of a process deals with actually creating a graphic representation of the process discovered (i.e BPMN of the process).
4. Assuring process model quality: Focus on ensuring the resulting process model meets different quality criteria (AVILA et al., 2020).

(DUMAS et al., 2018) presents two important roles for process discovery, which are defined as the *process analyst* and the *domain expert*. The first is responsible for gathering information about a given business process, and driving the modeling task (using modeling languages such as BPMN), under the leadership of the process owner. The latter, is an individual who has an expert knowledge of how a process or specific tasks are performed in details. This person is usually a participant of the process execution (e.g. A sales person responsible for executing sales quotation). The domain expert is crucial for the correct process discovery and modeling as the process analyst is usually a generalist individual with skills on BPMN modeling.

(DUMAS et al., 2018) also points out to three existing challenges in process discovery mainly associated to the *domain expert*. They are:

1. *fragmented process knowledge* - As a result of the specialization and division of labor it is not common that all activities of a processes are executed by the same person. It makes difficult for having holistic view of the business processes to be modelled.
2. *case level process knowledge* - In alignment with the first challenge, the second challenge is a consequence of the fact that the domain experts describe the tasks they conduct for one specific process instance but they might have problems responding how a process works in a general perspective.
3. *Unfamiliar with business process modeling languages* - The third challenge is the motivation why it has been defined two necessary roles for process modeling: The *process analyst* and the *domain expert*. The problem is that by lacking knowledge in BMPN, the domain expert also might fail at reading the model designed by the process analyst, making it difficult to validate the result of the process discovery.

Despite the challenges presented above, (DUMAS et al., 2018) illustrates three different approaches for making process discovery effective. These approaches are described below. Additionally, they are shown in the table 2.1.1, followed by a classification of its strengths and weakness:

- *Evidence-based*: based on documents, reports and observations;

- *Interview-based*: based on individual interviews with process expert;
- *Workshop-based*: based on interviews which considers all experts together at the same time.

Table 2.1: Table containing Process Discovery Methods

Method	**Approach**	**Strengths**	**Weakness**
Document analysis	Evidence-Based Discovery	• Structured information • Independent from stakeholders availability	. • Outdated material • Wrong level of abstraction
Observation	Evidence-Based Discovery	• Context-rich insight	. • Potentially intrusive • Stakeholders likely to behave differently • Only few cases and not all processes can be observed
Table continues on next page			

Automated discovery	Evidence-Based Discovery	• Extensive set of cases • Objective data	. • Potential issue with data quality and level of abstraction • Data may not be available or be available only in part • Data extraction and preparation is time-consuming
Interviews	Interview-Based Discovery	• Context-rich insights	. • Requires sparse time of stakeholders • Time-consuming: several iterations required before sign-off
Table continues on next page			

Workshops	Workshop-Based Discovery	• Context-rich insights • Direct resolution of conflicting views	. • Requires simultaneous availability of multiple stakeholder • Time-consuming: multiple sessions typically required

Source: The authors, 2021

Process Analysis

One of the advantages of BPM is gaining systematic insights into a process by the use of qualitative and quantitative measures. The qualitative method is of great importance for process analysis, however, the results from such analysis does not provide enough details to support on decision making (DUMAS et al., 2018). Whereas the quantitative measures of a process such as cycle time, waiting time and cost can provide valuable information to be used as basis for decision making over a process.

(DUMAS et al., 2018) presents four techniques for qualitative process analysis: Value-added analysis, Waste analysis, Stakeholder Analysis and Issue Documentation, and Root Cause Analysis. These techniques are briefly described below:

- *Value-added analysis*: As the name states, the tasks of a process are divided into steps and analyzed with the goal of identifying value adding activities, which are defined either as activities necessary for the business but do not provide positive outcomes to the client or if they have positive outcomes for the client, but does not add value to the business. The technique aims at eliminating the non-value adding activities.
- *Waste Analysis*: This technique can be seen as the opposite of the Value-added approach. It intends to find waste in business processes.

- *Stakeholder Analysis and Issue Documentation*: As part of the process analyst job, this technique aims at identifying issues, causes and unexpected events by gathering data from multiple sources and interviewing several stakeholders.
- *Root cause analysis*: Technique to investigate the root cause of undesired events or issues that could be spotted through processes analysis. An example is the German IKB bank that lost billions of dollars during the U.S. subprime mortgage crisis due to a wrong assessment of risk in their financial business processes (HERAVIZADEH; MENDLING; ROSEMANN, 2008).

For the quantitative process analysis, (DUMAS et al., 2018) presents three techniques: Flow Analysis, Queuing Analysis and Simulation. These techniques are briefly described below:

- *Flow Analysis*: Flow analysis aims at evaluating the overall performance of a process based on the performance of its tasks. As an example, it is possible to calculate the average cycle time of an end-to-end process by measuring the average cycle time of each task.
- *Queueing Analysis*: This technique is based on queuing theory, which is a field in mathematics. The goal is to analyze relevant parameters of a queue such as the expected length of the queue or the expected waiting time of an individual case in a queue.
- *Simulation*: It is the most straight-forward technique for quantitative process analysis, which implies simulation of the process execution to generate a large number of hypothetical instances of a process, executing these instances step-by-step, and recording each step in this execution. As result, it is obtained logs of the simulation, statistics of cycle times, average waiting times, and average resource utilization.

Process Redesign

The result from the process analysis phase may indicate a range of issues. In several cases, bottlenecks are identified, which slows down the execution of a business process. Often those issues are the motivation for process redesign.

Process redesign can be a complex subject, which involves methods, techniques and tools. In this chapter, it is briefly presented the use of the following three levels of abstractions to address processes redesign:

- *Methods*: It is a collection of problem-solving approaches ruled by principles and a common philosophy. Methods typically stretch out from the early analysis phase of a redesign project until the implementation of the proposed changes.
- *Techniques*: Well defined and standard procedures. Example of techniques used in process redesign are fishbone diagramming, Pareto analysis, and cognitive mapping. Supporting the redesign of a process, creativity techniques like brainstorming, SCAMPER, Six Thinking Hats, and Delphi are used.
- *Tools*: Information Technology (IT) systems used to support the execution of process redesign techniques. Most tools are in fact process modeling tools (i.e. BPMS, which supports the use of BPMN to capture a business process in a diagram).

Process Implementation

At this stage in the BPM lifecyle, only conceptual process models have been developed merely for discovery, analysis and documenting purpose. No technical implementation details are involved in the previous phases of the BPM cycle. For business process models to be interpreted and automatically executed by a software system, such as a BPMS, they must be systematically translated into executable process models. This procedure is defined as the *process implementation*.

(DUMAS et al., 2018) presents a five-step approach to incrementally transform a conceptual process model into an executable process. Below each of the steps is it summarized.

- *Identify the automation boundaries*: Based on the fact that not all processes can be automated, this first step starts by identifying which parts of a process can be controlled by a BPMS. This is done with the identification of three types of tasks aligned with the BPMN language: automated, manual, and user tasks. Automated tasks are executed by a BPMS or by an external system. Manual tasks are handled by process participants without the software support. User tasks stays in between automated and manual tasks.
- *Review Manual Tasks*: After the identification of the type for each process step, is it necessary to create a link of the manual tasks to the BPMS. This step is important due to the fact that if the task cannot be seen by the BPMS, it does not exist. The solution is either support manual tasks via technology or isolate these tasks and

automate the rest of the process

- *Complete the Process Model*: At this point in the process implementation, the automation boundaries and the manual tasks of the process being implemented are clear. This step ensures that the process model is complete. For that, two principles according to (DUMAS et al., 2018) are applied: (i) exceptions are the rule and (ii) no data implies no decisions and no task handoff. Those principles are applied due to the fact that conceptual process models neglect certain information, because modelers define it as irrelevant for the modeling purpose. However, information that is not relevant in a conceptual model may be highly relevant for a process model to be executed.
- *Bring the Process Model to an Adequate Granularity Level*: The granularity of a business process model in BPMN is usually an abstract model which may not be at the right level of granularity for implementation. A decomposition into finer-grained tasks is needed.
- *Specify Execution Properties*: As a result of the fourth step, it is obtained a to-be-executable process model, which consists of the right elements and the right level of granularity to be automated with a BPMS. However, this model is still independent of the technology it is going to be implemented. To make the model executable for a specific BPMS, it is necessary to specify in the last step how each element of the model is effectively implemented. These implementation details are called execution properties.

Process Monitoring

After finally being able to successfully implement a business process and going through all the BPM phases of identification, discovery, analysis, design and implementation, problems may still occur. It can happen that the process implemented does not meet expectations due to a variety of reasons such a business changes, regulations changes, system error and many others. This can result in operation issues, high processing times and exceptions. Additionally, problems with productive business processes usually affect the satisfactions of customers who would normally benefit from the value delivered by such a process. To deal with this situation, BPM lifecyle also counts with the *process monitoring* phase.

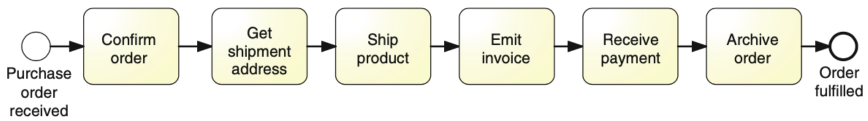
Two categories of process monitoring are presented by (DUMAS et al., 2018) and are briefly described below:

- *Offline Process*: Process monitoring focused on the analysis of historic process executions. The offline monitoring makes use o event logs which are categorized according to a timestamp. For this reason, offline process monitoring allows for analysis of cases completed during a particular period of time. The analysis can show poor performance, undesirable performance variations, and the conformance of the process according to certain rules or expected behaviors
- *Online Process Monitoring* The online monitoring of business processes aims at analysing business processes execution on-the-fly, which means the performance and conformance of a process is checked during runtime. This technique is mainly adopted by IT governance systems used for generating alarms or trigger counteractions whenever certain performance objectives or compliance rules are not fulfilled.

2.1.2 Process Aware Information Systems

Despite the recent popularity of RPA, which is the central subject of this work, process automation is not a new field. As it can be seen on the figure 2.7 located in the RPA section, business process automation has started on the 90s with the use of ERP systems. (DUMAS et al., 2018) describes process automation as an intent to automate any conceivable part of routine work that is contained within a business process, from simple operations that are part of a single process activity up to the automated coordination of entire, complex processes. Additionally, (DUMAS et al., 2018) exemplifies an automation example using a typical order-to-cash process as illustrated on the figure 2.4, which has been largely automated by the use of ERP systems since the 90s. In an automated order-to-cash process the seller of a product receives a purchase order, which is automatically dispatched to the ERP system of the warehouse and distribution department. The following step in the workflow is to check if the product is available in stock. If the product is not in stock, the suppliers are automatically contacted (also via IT systems) to manufacture the product. In case the product is available, the warehouse worker receives the PO and needs to manually retrieve the product from the warehouse. Finally, the sales department receives a notification that a new order needs to be confirmed. This is the behaviour of a typical PAISs, which usually implements automation of business processes or workflows. One example of industry, which successfully implements ERP systems for automation since the 90s, is the automotive industry. The car industry has automatated the entire value chain with end customers, distributors and suppliers via PAISs. The success of such industries is largely associated to their capacity to automate business processes and achieve global scale with reduced costs.

Figure 2.4: Figure illustrating a generic Order-to-Cash process



Source: (DUMAS et al., 2018)

Due to its characteristic of logging all information related to business process execution, PAISs play a crucial role in RPA and Process Mining. PAISs are capable of logging detailed information from a vast number of business processes. Usually PAISs implements logic behind financial, material management, HR and sales processes. As it is going to be explained later on the Process Mining section, the automatically discovery of business processes is largely dependent on such data.

The most common PAISs are presented below. The information is adapted from (DUMAS et al., 2018):

- *ERP Systems*: Provide essential business functionality, which is required across various industries. ERP systems support business processes in accounting and controlling, human resource management, and production management. The procure-to-pay and the order-to-cash process are the two most important processes that ERP systems fully cover.
- *Customer Relationship Management (CRM) systems*: Cover functionalities of marketing and sales processes that directly interact with customers. CRM systems help to document the interaction with individual customers through telephone, email, Internet portal, and personal encounters. In addition, CRM systems support sales and marketing activities related to products, pricing, distribution, and campaigning. CRM systems also acts as an extensive database that provides information on existing and prospective customers. The campaign-to-leads and lead-to-order are the most important business processes covered by CRM systems.
- *Supply Chain Management (SCM) systems*: These systems implements logistics op-

erations responsible for connecting suppliers and customers. SCM systems support the management of freight and transportation, inbound and outbound warehousing, storage and inventory, as well as corresponding planning and calculation processes. The order-to-delivery and return-to-refund processes are the most important processes implemented by SCM systems.

- *Product Lifecycle Management (PLM) systems*: PLM systems support the various processes of the engineering life-cycle of a product. These include the conception and design phase in which the product is specified, designed, and validated. The idea-to-launch, built-to-order, engineered-to-order, or assembled-to-order are the most important processes implemented using PLM systems.

2.2 Process Mining

This section presents an overview on Process Mining by defining important concepts and artifacts in which Process Mining depends on. Moreover, it also illustrates the most typical techniques used in the context of Process Mining.

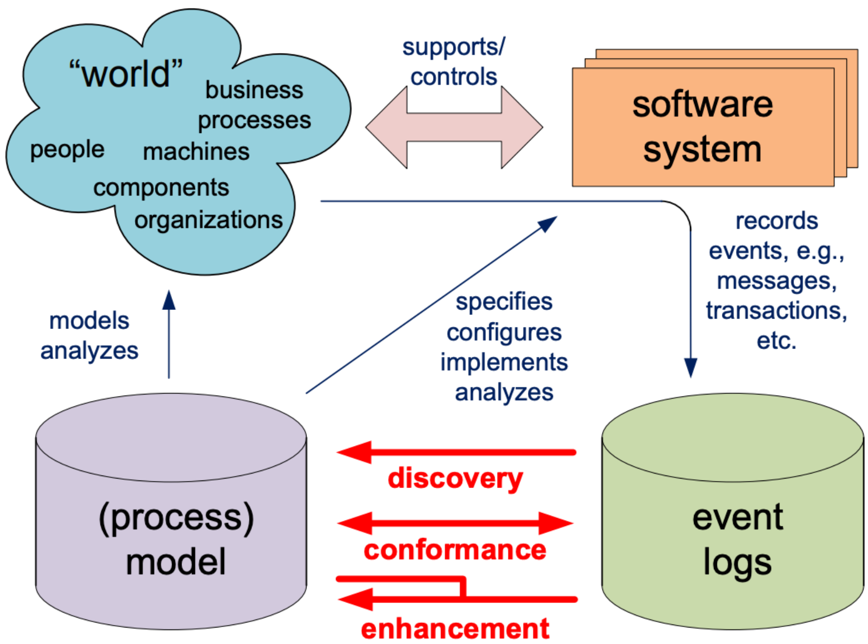
In the former section regarding BPM, it has been highlighted the importance of business processes as the backbone of any organization and enterprise value chain. It is believed the customer experience of any line of business relies on the quality of a service or product being delivered (BECKER; JAAKKOLA, 2020). Such results depends inherently on the business processes in place, and improvements on products and services require business processes changes, which are usually based on discovering of pitfalls and redesign of existing industry processes. Nevertheless, business processes adjustments are not simple do be done and can demand time, which is adversary to the challenging market speed.

In that context, Process Mining has emerged as a novel research field consisting of a combination of Data Science and BPM topics. (AALST, 2012) states that Process Mining aims at discover, monitor and improve real processes by extracting knowledge from event logs readily available in information systems. The knowledge from such event logs are by virtue of the PAIS also described in the BPM section of this work.

As also (AALST, 2012) describes, the starting point for Process Mining is an event log. Each event in such a log refers to an activity and it is related to a particular case of a process. The events belonging to a case are ordered and can be seen as one run of the process. Event logs may store additional information about events. Process Mining techniques use extra information such as the resource (i.e., person or device) executing or initiating the activity, the timestamp of the event, or data elements recorded with the event.

Process Mining can be used as a way to achieve three outcomes: One outcome of Process Mining is process *discovery*. A discovery technique is also based on an event log and can produce a business processes model without any human interaction or apriori information. Process discovery is the most prominent Process Mining technique. The second outcome of Process Mining is *conformance*. In this case, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third outcome of Process Mining is processes *enhancement*. In this last case, the goal is to extend or improve an existing process model with the use of information about the actual process execution recorded in some event log. Process enhancement aims at changing or extending the apriori model. With process enhancement it is possible to extend the model to show bottlenecks, service levels, and throughput times (outcome which would require a large effort if using traditional BPM process analysis techniques). Those concepts are illustrated in the figure 2.5 below:

Figure 2.5: Figure illustrating Process Mining concept



Source: (AALST, 2012)

2.2.1 Process Mining Techniques

This section introduces the most common techniques used in Process Mining for process discovery, conformance and enhancement.

Process Discovery

Different from the previous process discovery presented in the BPM section of this work, this section describes the discovery of a process using Process Mining instead of the manual approach used in traditional BPM. In Process Mining, business processes models are automatically discovered based on event logs collected from information systems.

The application of process discovery is motivated by a recurrent problem faced by organizations. Frequently, companies run their business processes in information systems (e.g ERP systems) without documenting or formalizing it. Often, when processes are documented, it is common the real process execution largely deviates from the business process model designed in BPMN systems. For this reason, process discovery has several applications for organizations. According to (AALST, 2012) process discovery may be used:

1. for discussing problems among stakeholders
2. for generating process improvement ideas
3. for model enhancement (e.g bottleneck analysis)
4. for configuring a BPMS

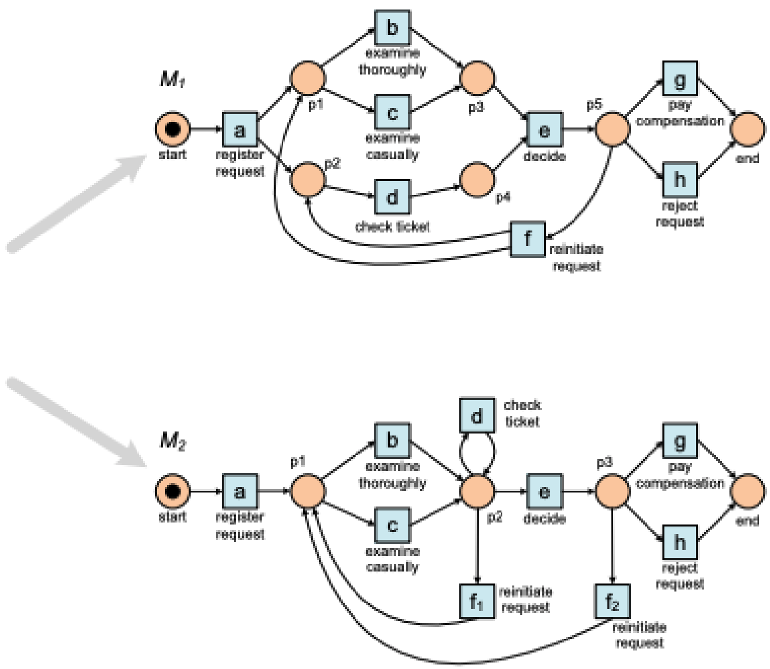
The α-algorithm is one of the most common methods used in the process discovery. The algorithm creates a process model based on a given event log. As illustrated in the image 2.6 from (AALST, 2012) , the α-algorithm produces model M_1 from the event log presented in the figure. This process model is represented as a Petri net, which consists of $places$ ($start, p_1, p_2, p_3, p_4, p_5,$ and end) and $transitions$ ($a, b, c, d, e, f, g,$ and h).Transitions may be connected to places and places may be connected to transitions. To represent the process flow, a Petri net uses the concept of $tokens$, which are distributed over places. The logic behind the workflow of a Petri net lies on the rule in which a transition is only enabled if each of its input places contains a token. As it can be observed in the figure 2.6, transition a is enabled in the initial marking of M_1, as a result of being the only input place containing a token, which is represented by the black dot.

Despite the ongoing popularity of Process Mining, the efforts for building algorithms to automate process discovery started in the mid-nineties. However, most of the classical techniques of process discovery present issues when dealing with concurrency, which is a common phenomenon in business processes. This problem is well addresses with the α-algorithm, which builds a Petri net based on dependencies identified in the event logs. Basically, the α-algorithm is a simple approach, which scans the event log for particular patterns. For example, if activity a is followed by b but b is never followed by a, then it is assumed that there is a causal dependency between a and b. The downside of the α-algorithm comes when dealing with complicated routing constructs.

For practical applications of process discovery, such as the Process Mining framework implemented by ProM, other algorithms are used. ProM's heuristic miner uses the algorithm described in (MEDEIROS; WEIJTERS; AALST, 2005), which builds a dependency graph based on the frequencies of activities and the number of times one activity is followed by another activity. The dependency graph reveals the backbone of the process model, which is used to discover the detailed split and join behavior of nodes.

Figure 2.6: Figure illustrating the process discovery with the use of system log and mining algorithm

#	trace
455	acdeh
191	abdeg
177	adceh
144	abdeh
111	acdeg
82	adceg
56	adbeh
47	acdefdbeh
38	adbeg
33	acdefbdeh
14	acdefbdeg
11	acdefdbeg
9	adcefcdeh
8	adcefdbeh
5	adcefbdeg
3	acdefbdefdbeg
2	adcefdbeg
2	adcefbdefbdeg
1	adcefdbefbdeh
1	adbefbdefdbeg
1	adcefdbefcdefdbeg
1391	



Source: (AALST, 2012)

Process Conformance

In addition to what has been presented in the former process discovery section, conformance checking uses both business process model and event logs as input. The business processes model may be resulted from manual design of the business process or an automatically generated model built by using Process Mining techniques. Conformance check aims at comparing the process model with the actual behaviour of the process captured through event logs. This is done by relating events in the log to activities in the model. As a result it is possible to compare the observed behavior in the event log and the modeled behavior.According to (AALST, 2012) process conformance may be used:

1. to check the quality of documented processes (asses whether they describe reality accurately),
2. to identify deviating cases and understand what they have in common,
3. to identify process fragments where most deviations occur,
4. for auditing purposes,
5. to judge the quality of a discovered process model,
6. to guide evolutionary process discovery algorithms (e.g., genetic algorithms need to continuously evaluate the quality of newly created models using conformance checking), and
7. as a starting point for model enhancement

Process Conformance can be implemented by using three techniques as described below:

- *Footprint*-based Technique - This technique uses a footprint of an event log and a footprint of a processes model, which shows a causal dependency between activities. In that sense, the technique aims at finding a disagreement among the footprints. For example, consider a footprint of an event log shows that a is sometimes followed by b and the opposite never occurs. In case the footprint of the corresponding model shows that a is never followed by b or that b is sometimes followed by a a conformance issue is identified by the technique.
- *Replay*-based Technique - As the name suggests, the second technique identifies a nonconformity in a process by simply replaying the event log on the corresponding process model. Such technique have two variations. The first, simply identifies

that at least one of the processes executions represented by the event log does not "fit" the process model, which is not a good quantitative indicator of how much the process model deviated from the real process execution. And the second variant, takes advantage of Petri net tokens in order to continuously replaying the event log on the model and “borrow tokens”, when deviations are spotted. In the end, the number of “borrowed tokens” and the number of tokens not consumed indicate the fitness level.

- *Alignment*-based Technique - The third and the most advanced technique aims at evaluating the conformance of a process by computing an optimal alignment between each trace in the log and the most similar behavior in the model.

According to (AALST, 2012) conformance check can be viewed from two angles: (a) the model does not capture the real behavior, which means the model is wrong and (b) reality deviates from the desired model, meaning in this case that the event log is wrong. At this point, it is possible to relate Process Mining conformance technique with the process monitoring phase described in the BPM section. Considering the difference that, in this case the process model could be automatically generated from the actual process execution (based on event logs).

Process Enhancement

The last Process Mining technique is usually of great interest for companies. Process enhancement allows extension or improvement of existing as-is process with the use of event logs. (AALST, 2012) A process model can be improved or corrected using the diagnostics provided by the alignment of model and log. Additionally, Process Mining techniques can also take advantage of the master data contained in the event log of process-aware systems (i.e. ERP systems). Information such as resources, timestamps, and case data (e.g. customer information, sales amount) can be retrieved. With the master data information (AALST, 2012) it is possible to analyze waiting times in-between activities (to identify the main bottleneck), reveal information about resources (to discover roles and groups of people frequently executing related activities) and analyze the decision points in the process model.

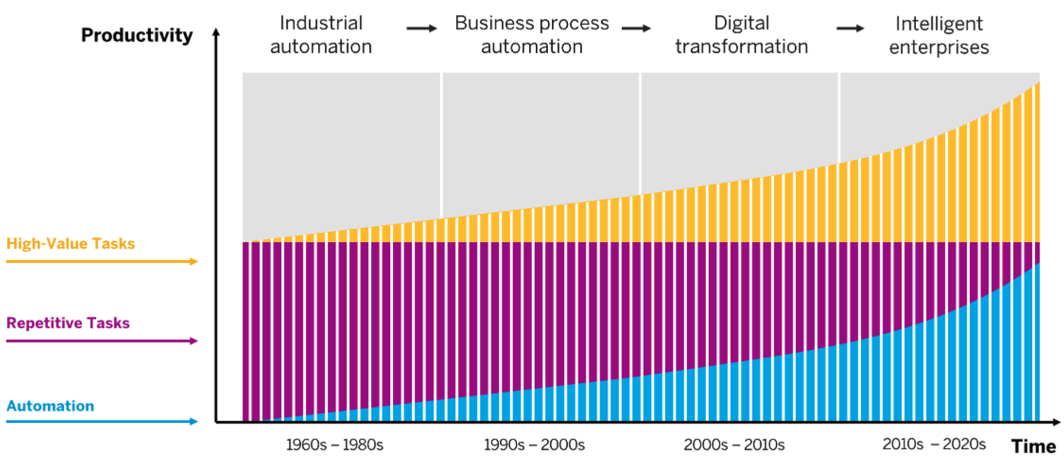
2.3 Robotic Process Automation

It is not new that companies are constantly aiming at thriving and prosper by increasing profit margin. For that, usually business leaders tend to focus on extending their operational limits to achieve the best results possible with the available resources.

What is considerable recent, is the strategy adopted by the enterprise for extending its operational limits to achieve better financial results. After decades of digital business transformation, companies are now starting to benefit from the knowledge and business information contained in IT systems, which are responsible for executing business transactions. A significant intellectual property of the enterprise is currently in posses of IT systems and not only humans. With the popularisation of AI and analytic solutions, IT systems are capable of providing business insights similar to humans. This is strategy is also referred as the Intelligent Enterprise (SRINIVASAN, 2016).

One of the main goals of the Intelligent Enterprise is to make use of technology to allow business experts to focus on high-value and strategic activities rather than the repetitive and tedious ones as seen in the image 2.7. In this direction, RPA emerges as a high potential player.

Figure 2.7: Figure illustrating evolution of automation in the industry



Source: (SAP, 2019)

RPA can be described as a technology, which aims at imitating a human worker with the goal of automating tasks in a fast and cost efficient manner. RPA is implemented by a computer software programmed to execute repetitive labour-intensive tasks (HOFMANN; SAMP; URBACH, 2020),. In more technical terms, the Institute for RPA (IRPA,

2019) defines RPA as the application of technology that allows employees in a company to configure SW-robots to capture and interpret existing applications for processing a transaction, manipulating data, triggering responses and communicating with other digital systems. It is important to note that in the context of RPA, robot does not mean a physical or mechanical machine. What RPA actually represents is a software-based solution, programmed to carry out procedures, processes or tasks on the repetitive way that are usually done by humans (SIBALIJA; JOVANOVIć; ĐURIć, 2019).

The most common use cases of RPA are sending emails, opening applications and copying and pasting information from one system to another (i.e. from Excel spreadsheets into ERP systems).

Another important fact is that on the image 2.7 it is possible to visualize Business Process Automation as a trend from 1990s to the 2000s. RPA differs vastly from Business Process Automation. The latter,(SIBALIJA; JOVANOVIć; ĐURIć, 2019) is a result of traditional BPM, which was a novelty in the 90's. The traditional BPM aim at process improvements by streamlining existing processes and removing inefficiencies (i.e. PAISs described in section 2.1.2 of this chapter). Therefore, this approach is based on creating or evolving systems and processes to increase efficiency. RPA is focused on enabling virtual workforce to do all the tedious, repetitive tasks. RPA does not optimize the process, instead it focus on making the processes execution faster, using software robots for performing process operations instead of human operators. And this is where most of the challenges of RPA are originated, which are going to be discussed in Section 2.3.1. In summary, as RPA does not improve existing business processes, process bottleneck and deviations. Process nonconformities must be addressed apriori to the automation. This is the point where Process Mining and BPM should support RPA.

As RPA is intended to carry out repetitive tasks, not all the processes are suitable for RPA implementations. According to (SYED et al., 2020), typical criteria for processes suitable for RPA are:

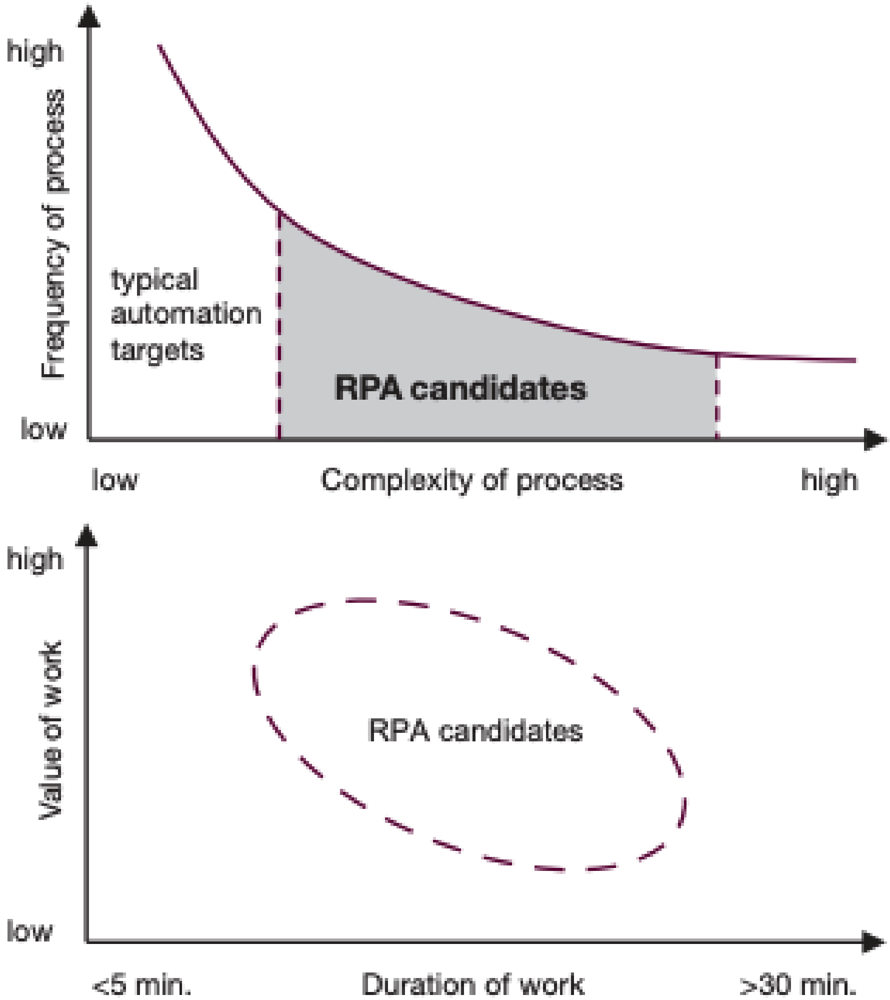
- *Highly rule-based*: the decision logic needs to be expressed in terms of business rules. RPA requires processes to contain well defined rules for every eventuality. Ambiguities should be removed apriori to the automation.
- *High volume*: one of the RPA greatest benefits comes from automating high volume tasks, which are usually time-demanding. Hence, sufficient transaction volumes help to maximise benefits from the implementation of software bots in an organisation.

- *Mature*: mature tasks are well consolidated processes, which are stable and people have good understanding of the activities involved in the processes.
- *Easy to achieve and show impact*: tasks performed within processes with the best return and simplest delivery. This make it easier to identify and highlight the business value for RPA.
- *Has digitised structured data input*: As RPA is SW-robot automation all input data must be digital and structured.
- *Highly manual*: Activities which do not require much human intervention, but are able to be automated.
- *Transactional*: RPA reduces the risk of transactional errors (e.g. incorrect data). For this reason, RPA is well suited for tasks dealing with transactional work. (i.e. typical tasks of ERP systems).
- *Standardised*: processes with a higher degree of standardisation are generally better candidates for the RPA implementation.
- *Low-levels of exception handling*: processes targeted for RPA should have the least amount of exceptions to deal. With a higher number of exceptions, the more RPA development, testing and optimisation will be delayed.
- *Highly repetitive*: automating repetitive tasks are likely to produce a return on investment faster.
- *Less complex processes*: Process complexity is directly related to bot implementation complexity. The less complex the processes being automated are, the lower are the risks for RPA to increase operating costs and possible cause business disruptions.
- *Well-documented*: the programming and testing of the bots are inherently dependent on the process documentation, which contains expected behaviour of the processes.
- *Interacts with many systems*: processes that need access to multiple system are good candidates for RPA. The manual effort for frequent access to different systems can be high and tends to result in human error.

In addition to those criteria defined above, a practical study conducted by (CAPGEMINI, 2016) showed RPA is ideal for replacing humans in repetitive tasks carried out 50-60 times a day, process list and file storage, ERP transactions, mass email generation, tasks performing conversion of data and periodic reporting.

As illustrated in the figure 2.8, (CAPGEMINI, 2016) study also categorized processes to be automated using RPA, considering processes frequency and complexity. Ideal to be automated are processes which are more complex and more frequent. There are also processes which should not be automated using RPA and those are the processes whose frequency is low and complexity is high. Besides process frequency and complexity, it has been analysed processes with a cycle time higher than 5 minutes and lower than 30 minutes as also good candidates for automation.

Figure 2.8: Figure illustrating criteria for candidate processes for RPA



Source: (CAPGEMINI, 2016)

2.3.1 Research challenges in RPA

This section summarizes the current RPA research challenges presented in (SYED et al., 2020). The challenges are categorized within different research topics, which are *benefits*, *readiness*, *capabilities*, *methodologies*, and *technologies*:

Table 2.2: Table containing RPA challenges

Challenge	**Research Theme**	**Description**
Support for benefit realisation	Benefits	The perceived benefits from RPA implementations in organizations can vary as it relies on organisational readiness for RPA, capabilities of the RPA technology to adopt, and implementation and delivery of an RPA solution. For this reason, the existence of systematic approach supporting benefit realisation of an RPA is of high importance to RPA implementation sucesses. Such systematic approaches rarely exists.
Comprehensive metrics for benefits	Benefits	The benefits deployed from RPA implementation are usually measured in terms of Total Cost of Ownership (TCO), which means reduction in time, cost, error, and human resources. Nonetheless, the benefits of RPA implementations might be much more valuable than TCO reduction if the capacity of human resources saved from repetitive tasks automated by RPA can be reallocated to more high-value strategies increasing enterprise productivity.
Models for organisational readiness assessments	Readiness	Companies require a readiness check framework in order to achieve strategic alignment by providing guidelines and tools to prepare for effective RPA implementations. Basically, it is important companies formally determine the potential opportunities and barriers for RPA deployment.
Mechanisms for infrastructure assessments	Readiness	It is also necessary for companies to be able to assess their existing technology infrastructure to support an RPA implementation. As a result, organisations may decide the conditions in which an RPA solution will best suit their needs

Models for organisational capabilities assessments	Capabilities	Organisations need an RPA capability assessment model to evaluate their organisational capabilities for automation and to assist in the roadmap for RPA programs.
Maximise analytical capabilities	Capabilities	RPA implementations still lack on AI and analytics functionalities. Further investigation on RPA is needed in order to allow the development of innovative solutions related to artificial and cognitive intelligence.
Methodological support for adoption	Methodologies	There is abundant information on the literature regarding the use of RPA from the strategic and managerial points of view, such as business drivers for RPA adoption and capabilities. Nonetheless, it is observed lack of synthesised recommendations and proposed approaches for RPA adoption with academic rigour.
Methodological support for implementation	Methodologies	There is a need for a methodology that focuses on the 'technical' considerations for large-scale RPA implementation. Although, there is an agreement on the use of Agile methodologies for the development of RPA, no consensus on what methodology should better address RPA implementations. bots.
Critical success factors	Methodologies	It is possible to encounter a high number of studies on RPA 'advisement' and RPA 'considerations'. However, it is not seen in the literature studies presenting a clear vision on what the critical success (or failure) factors are and how they may have different implications. (SYED et al., 2020) mentions a deeper understanding of RPA critical success factors can help firms to identify and better manage different elements to gain the best outcomes from RPA.

Socio-technical implications	Methodologies	Studies usually focus on presenting technical implication of RPA to organizations and lack of showing possible implications to IT/HR. It is necessary the development of organisational research that unveils the socio-technical implications of RPA..
Techniques for task selection	Technologies	Currently principles for selecting the candidate tasks for RPA are largely developed by specific RPA vendors and may be biased. It is needed a formal, systematic and evidence based techniques to determine the suitability of tasks for RPA.
Systematic design, development, and evolution	Technologies	The design of bots in RPA implementations are still largely a manual task, which can be tedious, inflexible, and error prone.Therefore, there is a need to develop and implement capabilities to systematically extract logical structures from user activities and transform these into algorithms for bot executions.
Seamless handling of exceptions	Technologies	RPA implementation relies strongly on a user interface, system interaction and change in business rules. For this reason, it is inevitable run-time exceptions might occur, which can lead to certain operational risks. It is clear the need for system-based, automated exception handling architectures and frameworks to maximise the of benefits RPA.
Techniques for managing scalability	Technologies	RPA implementations in a limited scope may perform well. On the other hand, in complex and large scenarios (typically common in Enterprise-wide adoption of RPA) scaling RPA solution might be a challenge. Innovative methods and techniques are needed to overcome the existing barriers to larger scale implementations.

Proactive monitoring and control	Technologies	At this moment bots are not capable of self-adapting to changes in business rules by self-monitoring its execution. For this reason, there is a need to develop new approaches to monitoring the runtime execution of bots and to proactively adapt to changes in business rules.

Source: (SYED et al., 2020)

2.4 Related Work

Based on the same motivation of this work, which is ruled by the scarce scientific material on RPA topic, (IVANCIC; VUGEC; VUKSIC, 2019) also presents a Systematic Literature Review (SLR) on RPA. The study mainly investigates how academic community defines RPA and to which extent it has been investigated in the literature in terms of the state, trends, and application of RPA. The paper also presents an overview of the RPA definitions and practical usage as well as benefits of its implementation in different industries. The results of (IVANCIC; VUGEC; VUKSIC, 2019) emphasises lack of theoretical studies on RPA, indicating that the area is still relatively new and that no theoretical frameworks have been formed.

In the context of Process Mining directly applied to RPA, (AALST, 2020) presents the Pareto Principle on Process Mining and RPA. The paper describes which combinations of Process Mining and RPA exists by presenting uses cases where Process Mining is applied in RPA implementations. The paper concluded that this combination can revitalize process management and address the traditional pitfalls of process modeling and process automation.

Table 2.3: Table containing related work details

Author	*Year*	*Title*	*Overview*
(IVANCIC; VUGEC; VUKSIC, 2019)	2019	Robotic Process Automation: Systematic Literature Review	Investigates academic research definitions of RPA through a SLR.
(AALST, 2020)	2020	On the Pareto Principle in Process Mining, Task Mining, and Robotic Process Automation	Overview on the combination of Process Mining and RPA. Presents use of Pareto Principle for RPA initiatives
(SALLET, 2021)	2021	Simplified Literature Review on the Applicability of Process Mining to RPA	Presents a Simplified Systematic Literature Review, which illustrates the main phases of RPA implementation and the respective use of Process Mining to RPA.

Source: The authors, 2021

3 SIMPLIFIED SYSTEMATIC LITERATURE REVIEW

A SLR is a methodological study, which aims at answering predefined research question within the scope of a research topic (KITCHENHAM, 2007). The SLR method includes rigorous and systematic activities such as collecting, summarising and classifying empirical evidences from existing studies in the research field being analysed. As a result of the rigorous methodology implemented by a SLR, the possibilities of biased and incoherent research results are significantly reduced.

The SSLR presented in this work is grounded in the following protocol proposed in (KITCHENHAM, 2007):

- Identifying the necessity of a research in the subject;
- Defining the research questions to lead the review;
- Composing a review protocol containing the research questions, inclusion and exclusion criteria, selection procedure, synthesis of the data;
- Evaluating the review protocol by an expert;
- Selecting primary studies;
- Assessing the primary studies;
- Extracting and synthesizing the data.

The SSLR in this work has identified the necessity of academic research of RPA and Process Mining, followed by a definition of research questions to conduct the research. Finally, Chapter 4 presents the selection of papers and extraction of data.

3.1 Protocol Review

This chapter explains the protocol applied during the construction of our SSLR. The review aims at presenting the applicability of Process Mining to RPA, as well as the Process Mining tools and techniques currently being adopted during the implementation of RPA.

3.2 Research Questions

The research questions aim at diminishing irrelevant results from the scope of the importance of Process Mining to RPA.

The questions were built mainly based on scientific papers, which highlighted the major research challenges in RPA, particularly the ones which mentions the use of Process Mining. As a result, a set of terms was created and applied to the research database presented on this chapter.

In order to test the relevance of the terms created, they were applied to the search engine of the selected research databases, where it was performed a random analysis from the first page of the search result. In addition, 5 papers were selected from each database to search if the terms had a significant recurrence within the content of the paper. At the end of the process, the following research questions were selected to guide the SSLR:

RQ1 – *What are the implementation phases of RPA?*
This question focus on establishing a terminology for the phases of RPA implementation in order to systematically answering the main topic of this research, which are described in the upcoming questions below. In addition to that, during the construction of the protocol it was identified inconsistent classification for naming the different implementation phases of RPA. Hence, this research question can help discovering the most frequent taxonomy for describing RPA implementation stages.

RQ2 – *In which implementation phases of RPA is Process Mining being applied?*
Based on the results from the first research question and the papers selected by the protocol, the second question aims at presenting the importance of Process Mining to RPA through the classification of which stages of RPA implementations Process Mining is being applied.

RQ3 – *What are the Process Mining tools and techniques being used in the implementation of RPA?*
Ultimately, the third research question goal is to illustrate how is Process Mining being used in RPA implementations.

3.3 Academic Databases

For our SSLR, it was selected four prestigious academic databases so as to achieve meaningful results on the importance of Process Mining to RPA. The academic databases selected are the following: IEEE Explorer; Digital Library; Springer Link; and Scopus.

3.3.1 Research Query

The research query aims at identifying the papers that discuss the applicability of Process Mining tools and techniques in the scope of RPA. Furthermore, the query also aims at reaching papers that describes the implementation phases of a RPA.

(rpa OR "robotic process automation") AND ("process mining" or "workflow mining") AND (implementation OR phase OR technique OR method OR tool)

3.4 Selection of Papers

The candidate papers were selected by applying a set of inclusion and exclusion criteria to the results obtained by the research query applied to the designated academic databases. This method is part of a systematic literature review and it is of great importance for selecting relevant results that could potentially answer the research questions defined in this work.

3.4.1 Inclusion Criteria

The inclusion criteria were defined with the intend to selecting the candidate papers based on their content and how related they are to the subject of the research. The considered inclusion (I) criteria (C) are the following:

IC-1 the paper directly describes Process Mining in the scope of RPA

IC-2 the paper mainly approaches RPA and only mentions Process Mining.

IC-3 the paper mainly approaches Process Mining and only mentions RPA.

IC-4 the paper presents a use case of Process Mining techniques or tools used in RPA.

The method used to verify the IC compliance was composed by the following steps, where the results of one step are the input to the next step:

1. The set of studies selected by the queries was analyzed in terms of title and author keywords of each piece. The focus was on expressions identical to the ones presented in the search query or related to them;
2. The abstracts were read to check if the pertinent expressions were related to the main scope of the study;
3. The conclusions were read to ensure the relation of the expressions to the scope of the study.

3.4.2 Exclusion Criteria

The exclusion (E) criteria (C) were determined to exclude irrelevant papers in terms of format, publication details and access. The content of the papers is not taken into account. The selected EC are:

EC-1 the paper is not written in English.

EC-2 the paper has less than 4 pages, considered not a full paper.

EC-3 articles without full access

EC-4 articles that are actually a book chapter

EC-5 the result from the query is not a paper

EC-6 The paper has not been published prior to 2017

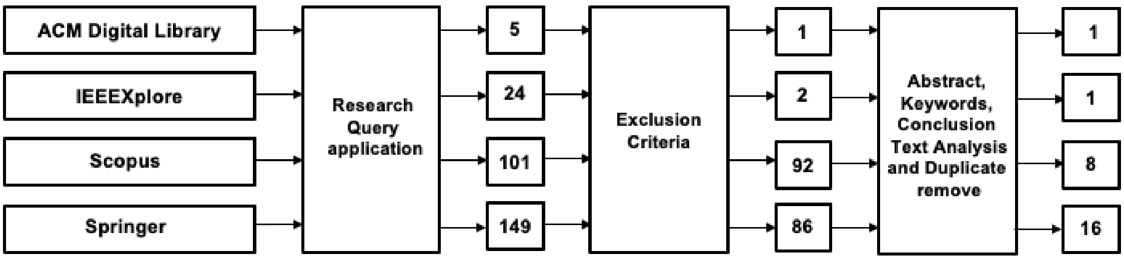
The EC were applied by filtering the set of results from the search engines of the academic libraries. EC-1, EC3, EC4, and EC-5 were applied preexisting filters that the academic databases provide. The EC-2 was applied by analyzing the publication and format of the paper.

4 A CLASSIFIER FOR RPA IMPLEMENTATION PHASES AND PROCESS MINING USAGE IN RPA

This chapter presents the results of our SSLR, which aims at answering the three research questions described in the chapter 3 of this work.

It also presents a classifier for the RPA implementation phases along with a mapping of Process Mining applicability to the respective RPA stages identified through the SSLR. The figure 4.1 shows the results obtained by applying the research query and EC followed by a manual selection of papers based on the IC, where we analyzed the paper's abstract, keywords, and conclusion. In addition, the resulting selected papers after text analysis and removal of duplicated articles is illustrated on the table 4.1.

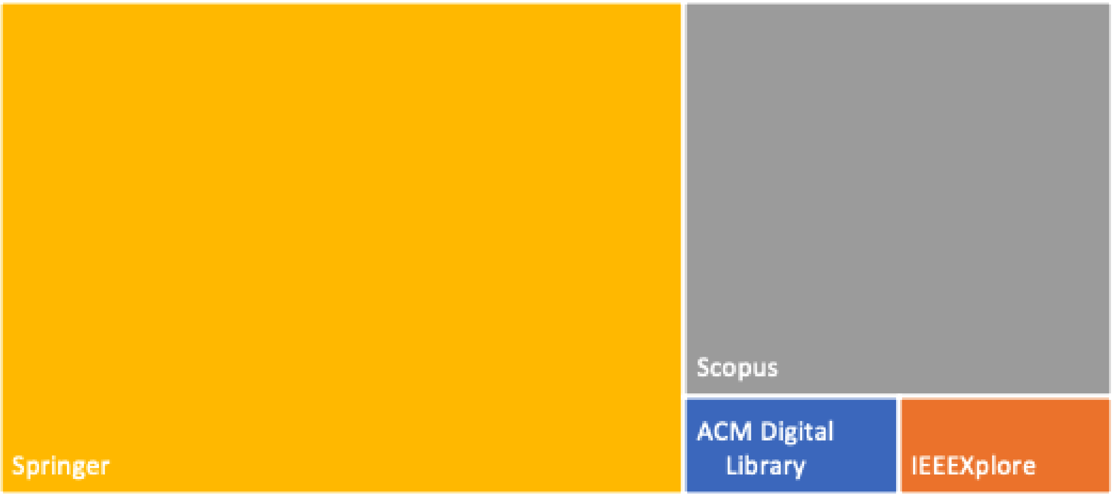
Figure 4.1: Illustration of systematic paper selection



Source: the authors, 2021

As it can be seen in the figure 4.2, the majority of articles used in the SSLR originated from Scopus and Springer academic databases. In fact, the research query for ACM Digital Library and IEEEXplore had significantly less successful hits. As Scopus tends to index other academic databases it has also been found a high number of duplicated articles in the search engine results of Scopus. This is also one the reasons for the reduced number of papers actually evaluated. Another reason for that, is the fact that the inclusion criteria proposed by our SSLR clearly states the exclusive interest on papers which describes both RPA and Process Mining together. There is a large number of papers, which analyses RPA or Process Mining in a independent way and, for that reason, were excluded from our SSLR.

Figure 4.2: Selection of papers by academic database



Source: the authors, 2021

Table 4.1: List of all selected papers

Paper	*Citation key*	*Title*
[1]	(EGGER et al., 2020)	Bot Log Mining: Using Logs from Robotic Process Automation for Process Mining
[2]	(KÖNIG et al., 2020)	Integrating Robotic Process Automation into Business Process Management
[3]	(HERM et al., 2020)	A Consolidated Framework for Implementing Robotic Process Automation Projects
[4]	(JIMENEZ-RAMIREZ et al., 2019)	A Method to Improve the Early Stages of the Robotic Process Automation Lifecycle
[5]	(CHAKRABORTI et al., 2020)	From Robotic Process Automation to Intelligent Process Automation
[6]	(AGOSTINELLI; MARRELLA; MECELLA, 2019)	Research Challenges for Intelligent Robotic Process Automation
[7]	(WELLMANN et al., 2020)	A Framework to Evaluate the Viability of Robotic Process Automation for Business Process Activities

[8]	(GAO et al., 2019)	Automated Robotic Process Automation: A Self-Learning Approach
[9]	(CABELLO; ESCALONA; ENRÍQUEZ, 2020)	Beyond the Hype: RPA Horizon for Robot-Human Interaction
[10]	(AGOSTINELLI et al., 2020)	Automated Generation of Executable RPA Scripts from User Interface Logs
[11]	(PARK; AALST, 2020)	A General Framework for Action-Oriented Process Mining
[12]	(LÓPEZ-CARNICER; VALLE; ENRÍQUEZ, 2020)	Towards an OpenSource Logger for the Analysis of RPA Projects
[13]	(GUPTA et al., 2020)	Analyzing Comments in Ticket Resolution to Capture Underlying Process Interactions
[14]	(PAUWELS; CALDERS, 2020)	Bayesian Network Based Predictions of Business Processes
[15]	(LENO et al., 2020b)	Robotic Process Mining: Vision and Challenges
[16]	(AUGUSTO et al., 2019)	Split miner: automated discovery of accurate and simple business process models from event logs
[17]	(CERNAT; STAICU; STEFANESCU, 2020)	Towards automated testing of RPA implementations
[18]	(ENRíQUEZ et al., 2020)	Robotic Process Automation: A Scientific and Industrial Systematic Mapping Study
[19]	(LENO et al., 2019)	Action Logger: Enabling Process Mining for Robotic Process Automation
[20]	(LENO et al., 2019)	Multi-Perspective Process Model Discovery for Robotic Process Automation
[21]	(WANNER et al., 2019)	Process Selection in RPA Projects – Towards a Quantifiable Method of Decision Making

[22]	(AALST, 2020)	On the Pareto Principle in Process Mining, Task Mining, and Robotic Process Automation
[23]	(LINN; ZIMMERMANN; WERTH, 2018)	Desktop Activity Mining - A new level of detail in mining business processes
[24]	(LENO et al., 2020a)	Identifying candidate routines for Robotic Process Automation from unsegmented UI logs
[25]	(RIZK et al., 2020)	A Conversational Digital Assistant for Intelligent Process Automation
[26]	(FISCHER et al., 2021)	On the composition of the long tail of business processes: Implications from a Process Mining study

Source: The authors, 2021

4.1 Overview of Findings

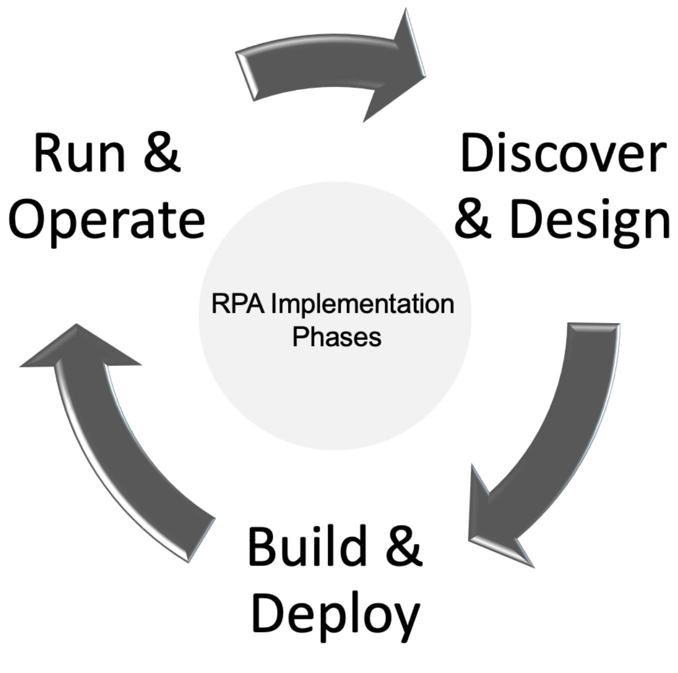
This section is divided in three parts. The first part illustrates a model created in this work to classify the different phases of an RPA implementation by mapping the taxonomy used in the papers of our SSLR into the proposed model. The second part, makes use of the classification of RPA implementation phases proposed in the first research question of our SSLR to point out which are the phases of RPA projects in which Process Mining is applied. Finally, the last section presents a summary of the Process Mining techniques and tools that were encountered in the SSRL, which evidences the use of Process Mining in RPA.

4.2 Research Question 1: Implementation phases of RPA

Motivated by the diversity of taxonomy used in the literature to describe the different implementation stages of RPA, we decided to build a nomenclature on a higher tier to better classify RPA implementation phases and, consequently, be able to systematically answer the research questions of the SSLR. Most articles, which describes RPA implementation, use different terminology for naming the activities and stages involved in RPA projects. We proposed a model and presented a possible correspondence to the BPM lifecyle. In summary, it has been observed during the research that RPA constructions typically incorporates three main phases: *Discover & Design*, *Build & Deploy* and *Run & Operate*. The BPM phases of process identification, process discovery, process analysis and process redesign were associated to the Discover & Design phase of the proposed model due its similarities in activities. The same way, the process implementation phase of BPM has been linked to the Build & Deploy phase of the RPA implementation phases model. Finally the Run & Operate phase has been related to the process monitoring of the BPM cycle.

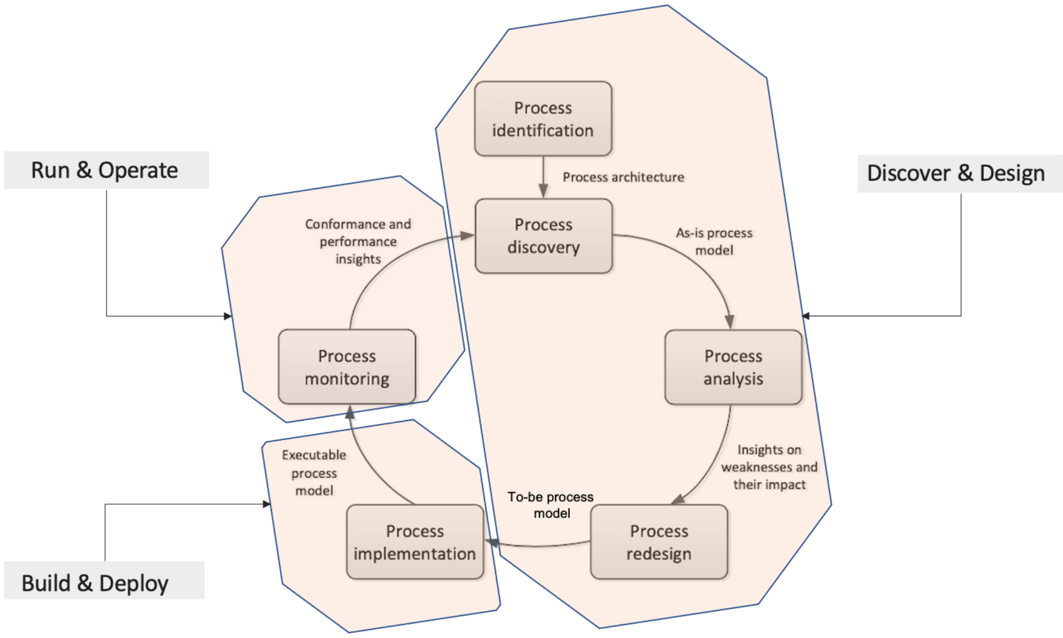
The model proposed and its relationship to the BPM lifecyle is illustrated respectively in the images 4.3 and 4.4. Further on this section, each of the proposed RPA phases are explained based on the evidences encountered through the conducted research.

Figure 4.3: Proposed model for RPA implementation phases



Source: The authors, 2021

Figure 4.4: Figure illustrating relationship between BPM lifecycle and the proposed model for RPA implementation phases



Source: Adapted from (DUMAS et al., 2018)

4.2.1 Discover & Design

The majority of articles, which presents a case study of a practical RPA project, describes that the initial phase of an RPA implementation mainly consists of identifying candidate routines to be automated [4]. Besides the discovery of the process to be automated, the early stages of RPA implementation also involves design activities such as the specification of actions and the definition of the data flow to be developed, which often incorporates record of events that happen on the User Interface (UI) of the user's computer system [6]. The set of activities performed in the initial stages of RPA implementation has been classified as the *Discover & Design* phase.

Due to the nature of activities present in the Discover & Design phase, this initial stage of RPA implementation is frequently characterised as the most challenging part of RPA. Discovering which process to automate can be an arduous task, it may involve manual analysis of the activities performed by the process's stakeholders through the observation of their actions, which might not offer scalability and efficiency when a high number of processes has to be considered, or through analysis of process-related documentation, which has the risk of containing outdated information [5]. Such activities do not differ much from the methods and techniques presented in the process identification, discovery, analysis and (re)design of the BPM lifecycle. For this reason, the proposed model in figure 4.4 illustrated the association of the Discover & Design phase to the initial stages of the BPM cycle. As RPA is a result of a software implementation of a process, defining which technologies and tools are going to be used for building the designed RPA is also of great relevance when planning the construction of an RPA solution [3].

Predominantly, the processes to be automated, the design of the elected process, typically in BPMN, and the technologies to be used are the result of the Discover & Design phase of an RPA implementation. It is important to note that besides the challenge of identifying which processes to automate, it is also necessary that the BPMN resulted from the process discovery represents a standardised version of the processes. Essentially, RPA itself does not handle exceptions, and issues encountered during this phase must be solved through process analysis and redesign.

4.2.2 Build & Deploy

Once the candidate processes to be automated have been selected, appropriately designed and the tools and technologies to be used in the implementation have been elected, it is started the RPA phase named *Build & Deploy*. This phase consists mainly of highly technical activities such as configuration and development of code, which relates to the BPM phase cycle named *process implementation*. Typically, in RPA projects this phase is used for building a proof-of-concept for the purpose of testing the elected process or a simple instance of the process, which provides a good opportunity for assessing technical and financial feasibility of RPA technology [3]. On the other hand, paper [4] points out RPA implementations usually lack of productive system, and differently from what occurs in traditional software development lifecycle, where testing precedes deployment, RPA is characterized by testing directly in the production environment along with the deployment of the SW robot solution.

The Build & Deploy phase of RPA projects aims at delivering a ready-to-run SW robot automation of a certain business process.

4.2.3 Run & Operate

At the final phase of an RPA implementation, a software robot is deployed on its respective execution environment. As any productive software solution, RPA requires a series of controlling, monitoring and performance evaluation activities in order to ensure the SW robot meets the expectations of the designed automation and no business disruptions are caused due to a bot malfunctioning. This stage is classified as the *Run & Operate* phase. In addition to the basic monitoring of business processes execution, similar to the process monitoring from BPM lifecyle, and bot performance monitoring, [5] evidences that the construction of a Center of Excellence, containing RPA experts and resources capable of giving maintenance to the running RPA solution, is essential to support continuous changes and adjustments in RPA implementations. The *Run & Operate* is the RPA phase responsible for the maintenance of the SW bot and the entire RPA implementation by ensuring a long-term service of RPA integrated in the production processes and cooperation between humans and machines.

Table 4.2: Papers containing evidences for the RQ1

Paper Ref	*RPA Phases cited*	*Mapping to proposed taxonomy*
[2]	Design and Analysis	Discover & Design
	Configuration	Build & Deploy
	Enactment	Run & Operate
	Evaluation	Run & Operate

[3]	Identification of Automation Need	Discover & Design
	Alignment with Business Strategy	Discover & Design
	Screening of Different (RPA) Technologies	Discover & Design
	Processes Selection	Discover & Design
	RPA Software Selection	Discover & Design
	Proof of Concept Implementation	Build & Deploy
	Evaluation of Business Case	Build & Deploy
	RPA Rollout	Build & Deploy
	Adaptation and Scaling of RPA Services	Run & Operate
	Setting up Center of Excellence	Run & Operate
	RPA Support Processes	Run & Operate
[4]	Analysis	Discover & Design
	Design	Discover & Design
	Development	Build & Deploy
	Testing	Build & Deploy
	Deployment	Build & Deploy
	Operation and Maintenance	Run & Operate
[6],	Determine process to automate	Discover & Design
[10]	Model the selected process	Discover & Design
	Record UI events	Build & Deploy
	Develop	Build & Deploy
	Deploy	Build & Deploy
	Monitor the performance	Run & Operate
	Maintenance	Run & Operate
[7]	Evaluate RPA automation candidates	Discover & Design
[17]	Test	Build & Deploy
[18]	Analysis	Discover & Design
	Design	Discover & Design
	Construct	Build & Deploy
	Deployment	Build & Deploy
	Control and Monitoring	Run & Operate
	Evaluation and Performance	Run & Operate

Source: The authors, 2021

4.3 Research Question 2: Implementation phases of RPA, in which Process Mining is applied

This section aims at answering the RQ2, which focus on presenting the applicability of Process Mining to RPA by illustrating the usage of Process Mining among the three different phases of RPA implementation.

The RQ1 presented in section 4.2 along with the research challenges in RPA described in the section 2.3.1 of this work, has illustrated that the early stages of RPA implementation are time-consuming, often relying on the study of process documentation, which is typically incomplete, inaccurate or differs from reality. As a result, the Discover & Design phase of RPA project is usually the most challenging period of SW robot building. For that reason, not surprisingly the SSLR has shown that most of the Process Mining efforts, currently documented in the literature, aims at improving the early stages of the RPA lifecycle [4].

Besides that, is has been observed also significant efforts for applying Process Mining for Run & Operate phase of RPA, with the motivation of providing operational support with use of predictive process monitoring to ensure continuous process management and improvement [11].

The least frequent use of Process Mining in RPA stages is in the Build & Deploy phase. This phase, as presented in section 4.2.2, consists of mostly technical activities (i.e. code development and configuration), which are highly dependent on human resources. However, studies such as the ones presented in [5], [8], [10] demonstrates effort in using Process Mining also for the Build & Deploy phase by combining AI and Process Mining techniques. Study [5] presents the concept of Intelligent Process Automation (CHAKRABORTI et al., 2020), which consists of traditional RPA technologies combined with AI and Process Mining to automate complex tasks which require decision making, insights and analysis. As an example, [10] describes a method that build automated RPA scripts through UI event logs, reducing the need of qualified human to configure or code a SW bot. Table 4.3 lists the papers used to answer RQ2 followed by their respective application in RPA.

Table 4.3: Papers containing evidences for the RQ2

Paper Ref	*RPA Phases with Process Mining applications*
[1]	Run & Operate
[2]	Run & Operate
[3]	Discover & Design
[4]	Discover & Design
[5]	Discover & Design + + Build & Deploy + Run & Operate
[6]	Discover & Design
[7]	Discover & Design
[8]	Discover & Design + + Build & Deploy
[9]	Run & Operate
[10]	Discover & Design + + Build & Deploy
[11]	Run & Operate
[12]	Discover & Design
[13]	Discover & Design
[14]	Discover & Design
[15]	Discover & Design
[16]	Discover & Design
[20]	Discover & Design
[21]	Discover & Design
[22]	Discover & Design
[23]	Discover & Design + Build & Deploy
[24]	Discover & Design
[25]	Build & Deploy + Run & Operate
[26]	Discover & Design

Source: The authors, 2021

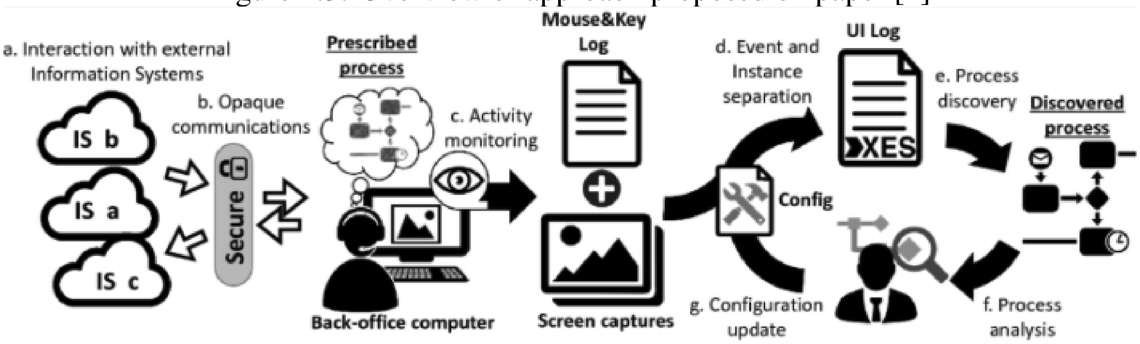
4.4 Research Question 3: Process Mining tools and techniques used in RPA Implementations

The last research question aims at presenting details of Process Mining tools and techniques applied to each of the three previously defined phases of RPA implementation.

Process Mining Techniques used in the Discover & Design phase of RPA

The first phase of an RPA implementation, as described in the RQ1, consists mostly in discovering ideal candidate processes to be automated, which is highly challenging stage of RPA projects. For this reason, as also demonstrates the results in RQ2, the majority of papers provide methods and techniques for the Discover & Design phase. Paper [4] describes and evaluates an approach for the early stages of an RPA project. The approach focus in gathering knowledge of the activities performed in IT systems by a back-office staff. It starts by monitoring the activities performed by the back-office staff in a non-invasive manner with the use of a screen-mouse-key-logger. The log obtained is transformed into an UI log through image-analysis techniques and then turned into a process model by the use of process discovery algorithms. This technique can be classified as Desktop Activity Mining (DAM), which is introduced by the paper [23]. Image 4.5 summarizes the approach described in [4].

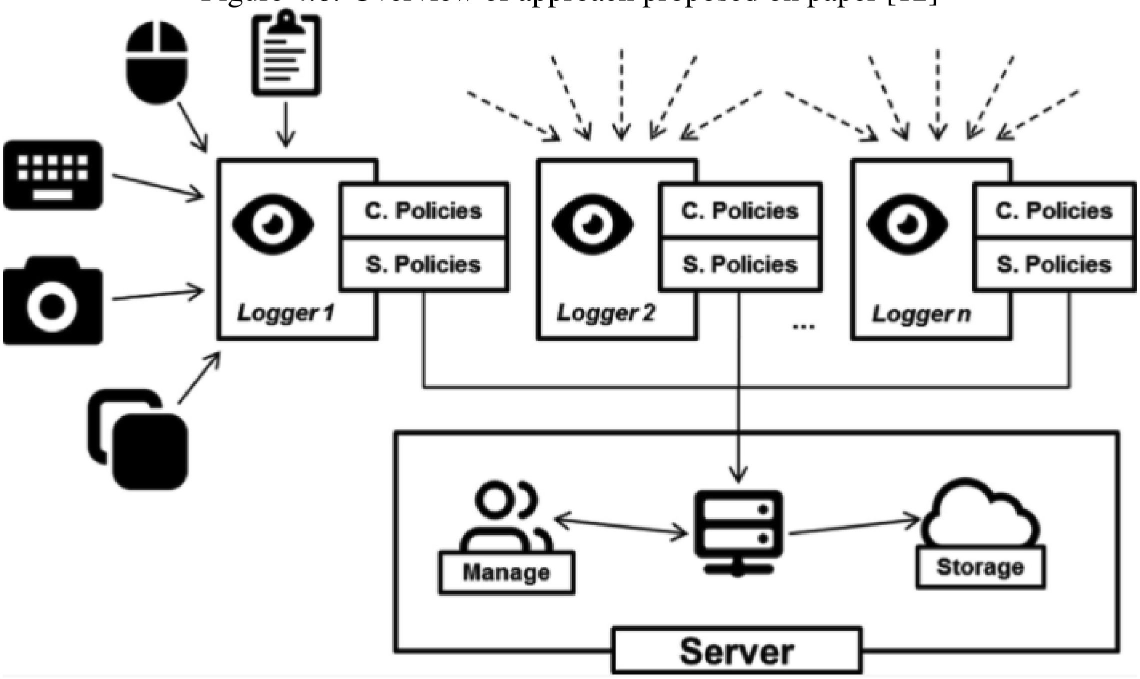
Figure 4.5: Overview of approach proposed on paper [4]



Source: (JIMENEZ-RAMIREZ et al., 2019)

Paper [12] presents a multi-platform Open Source UI logger which generates UI logs in a standard format. This approach collects information from all the computers it is running on, and sends it to a central server for processing. After processed, the collected information allows the creation of enriched UI logs, which may be used for process analysis, machine learning training and eventually the creation of RPA robots. Figure 4.6 illustrated the approach.

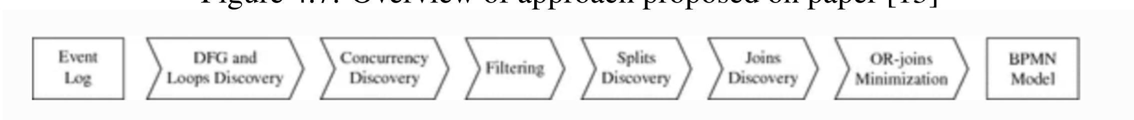
Figure 4.6: Overview of approach proposed on paper [12]



Source: (LÓPEZ-CARNICER; VALLE; ENRÍQUEZ, 2020)

The literature shows that automated process discovery methods usually presents deficiencies when applied to real-life logs by resulting in spaghetti-like processes models. Such models typically poorly fits the respective event log or over-generalize the processes by lacking of relevant details. [13] presents an automated process discovery method, called *Split Miner*. Split Miner incorporates an approach to filter the flowchart originated from event log, aiming at identifying combinations of split gateways that accurately capture the concurrency, conflict and causal relations in the flowchart. Split Miner is considered to be the first automated process discovery method that is guaranteed to produce deadlock-free process models with concurrency. The image 4.7 illustrates the approach.

Figure 4.7: Overview of approach proposed on paper [13]



Source: (GUPTA et al., 2020)

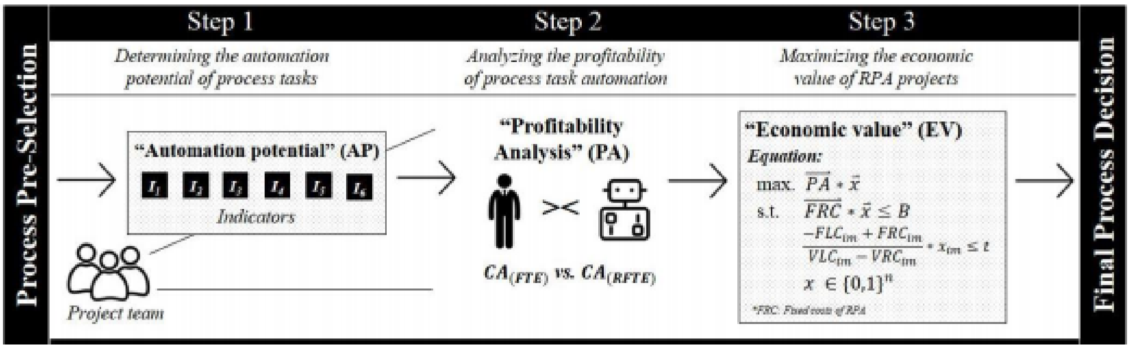
The paper [19] also presents a method for the Discover & Design phase, which consists of a tool named Action Logger used for automatic recording of user interactions with IT systems. The logs recorded by Action Logger from the UI are developed to be imported into Process Mining tools, such as Apromore (ROSA et al., 2011).

Paper [8], presents an approach based on the capture of user behaviour, which automatically detects high-level RPA-rules. In the first step, tasks are identified by observing the user's interactions with the systems, on a collection of system forms that are defined. On the second step it deduces rules by learning relations between the different tasks performed. This approach also support RPA implementation in the Build & Deploy phase with the automatic generation of instructions that can be executed by RPA tools. Such rules are defined as "if ... then ..." statements Finally, it applies the rules by instantiating RPA on the basis of the deduced rules. The "then" part of the rule represents actions that can be automatically executed by the RPA solution.

Paper [20] presents an automated discovery approach for data-aware declarative models, which means models that are open and offer more possibilities for execution. Differently from the procedural models (i.e. based on event logs from ERP systems), which explicitly specifies the flow of the interactions among process activities. A declarative model describes a set of constraints that must be satisfied throughout the process execution. The approach allows for automated discovery of multi-perspective declarative process models and it is able to discover conditions involving arbitrary data attributes.

With the use of Process Mining, paper [21] address a well known challenge of RPA for process selection determination, as described in Chapter 2 of this work. The approach aims at quantifiable measuring the value of RPA project implementation by automatically evaluating which RPA activities should be selected for automation in order to maximise RPA's return on investment. Based on defined indicators, the method initially determines the automation potential of processes, followed by an analysis of the profitability of process task automation, and finally maximise the value on RPA projects. As a result, the method returns a list with quantified indicators and recommendations to support the subsequent decision-making for RPA processes selection. The method is illustrated in figure 4.10.

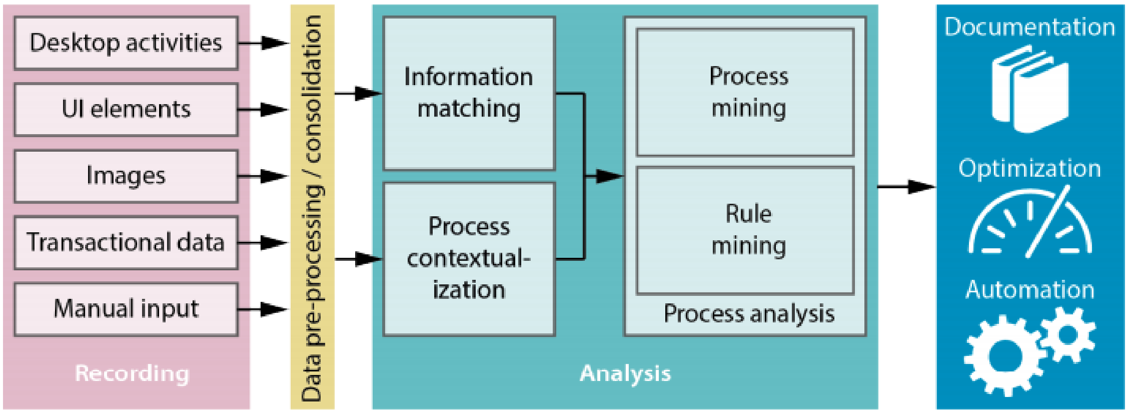
Figure 4.8: Overview of approach proposed on paper [21]



Source: (WANNER et al., 2019)

During the SSLR, Desktop Activity Mining (DAM) has showed to be the most typical technique applied in Process Mining methods for RPA. Paper [23] presents DAM as an approach to mine detailed process activity data. The idea is to make use of detailed desktop activities of all users performing an office process and consolidate the process variations with Process Mining techniques to discover the resulting process model to be automated. It can be observed in papers [4], [8], [10], [12], [19] and [24], that the DAM approach is to a certain extend, also applied. The DAM method concentrates its efforts on recording desktop activities, UI elements, images, transactional data and manual input from IT systems, where users are involved in business processes execution. It is followed by an analysis phase of the event log generated through the recording of activities. The final result is a process model and documentation, which can further be used for redesign and automation of the business processes being mapped. It is important to notice that Desktop Activity Mining and Process Mining are different approaches, which complements each other for data-driven process discovery and documentation. Process Mining is only used after the UI log is generated via the recording of the business processes activities. The image 4.11 , illustrates this approach.

Figure 4.9: Overview of approach proposed on paper [23]



Source: (LINN; ZIMMERMANN; WERTH, 2018)

The paper [24] describes an approach to automatically identify routines from unsegmented UI logs, which means the approach can discover processes resulting from event logs where a set of traces of a task does not contain one or more routines. When the log is segmented, the identification of candidate routines is done by discovering frequent sequential patterns from a collection of sequences. The approach for unsegmented logs, starts by decomposing the UI log into segments corresponding to paths within the connected components of a Control-Flow Graph (generated from the UI log). Finally once the log is segmented, a pattern mining technique is used to extract frequent patterns. The patterns are then ranked according to four quality criteria: frequency, length, coverage, and cohesion.

Process Mining Techniques used in the Build & Deploy phase of RPA

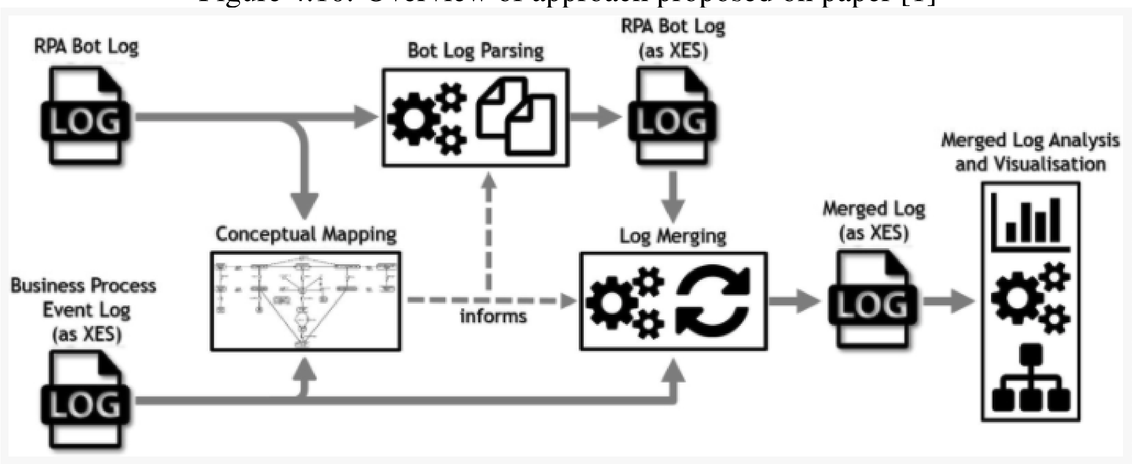
For the Build & Deploy phase, a smaller amount of techniques were identified. This could be due to the nature of activities during this phase, which are technical activities such as configuring and coding that requires skilled human experts. However, it has been also presented a few promising Process Mining techniques that could unleash the potential of automating RPA construction. As an example, paper [10] presents a cross-platform tool that allows User UI logs to automatically generate executable RPA scripts.

Moreover, paper [17] describes a model towards the automation of RPA testing, however, it does not present details. It only presents the concept and viability of applying Model-based testing (MBT), that is an approach that uses a model of the system under test to automatically generate test cases, which can be applied in RPA testing (the model of the business processes is the basis for the automation). Considering the nature of RPA implementations, creating tests based on the processes model to test RPA can be an effective method indeed. The paper concludes that the approach must be replicated using RPA software providers such as Automation Anywhere and Blue Prism. Additionally, it reinforces the need of developing research in the topic of automation of RPA tests in the academia.

Process Mining Techniques used in the Run & Operate phase of RPA

There is also a significant number of papers which showed efforts to support RPA with Process Mining by exploring the final stage of RPA, which has been defined in this work as the Run & Operate phase. [1] has proposed an approach to help improving existing RPA implementations and exploring opportunities for bot and process redesign. In this approach it is shown that historical data from RPA-enabled processes in the form of bot logs or process logs can be utilized for RPA enhancements. The approach proposed merges bot logs with process logs for Process Mining. The image 4.10 illustrates this approach.

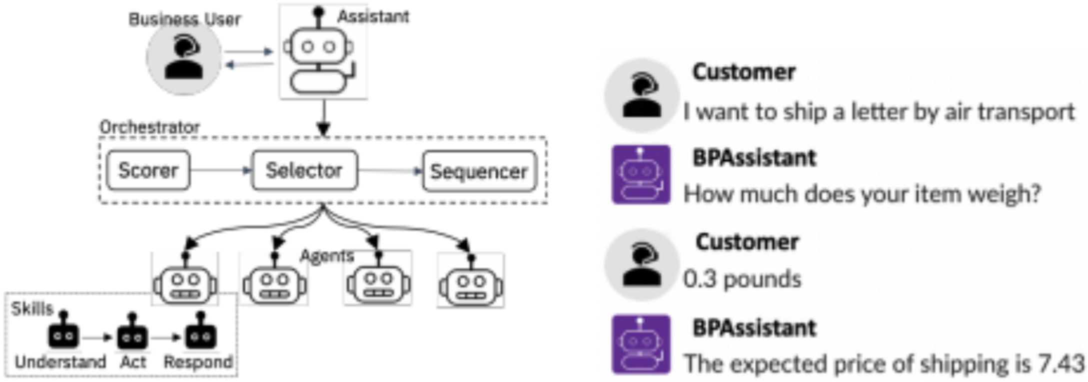
Figure 4.10: Overview of approach proposed on paper [1]



Source: (EGGER et al., 2020)

As one of the few techniques presented for the Run & Operate phase of RPA implementations, paper [25] introduces a framework that combines RPA with conversational agents (also named as chatbots). The framework aims at creating an interactive business process automation solution, with the use of natural language techniques and Process Mining. The framework relies on multi-agent orchestration, where conversational agents are composed from RPA. The resulting assistant allows business users to monitor and customize their business process automation solutions through the use of natural language. This framework can be considered to provide support in the Build & Deploy phase of RPA, since it can result in RPA configuration adjustments according to the interaction of business users and chatbots. The image 4.11 illustrates this approach.

Figure 4.11: Overview of approach proposed on paper [25]



Source: (RIZK et al., 2020)

5 CONCLUSION

This work conducted a SSRL to present the applicability of Process Mining to RPA. Besides that, it presented a classifier for RPA implementation phases along with the respective Process Mining uses on the RPA distinct phases.

Throughout the paper's analysis, it was concluded that the taxonomy for defining implementation phases of RPA is diverse. Nonetheless, the different terms used to narrate each of the RPA phases contains equivalent semantics when assessing the details of the RPA implementation activities described by the paper's authors. For this reason, this work proposed a model for grouping RPA implementation stages, which consists of three main phases: Discover & Design, Realize & Deploy and Run & Operate. The proposed model has been created based on the literature and has not yet been tested.

The analysis also revealed that the Discover & Design phase of RPA implementations is the most time-demanding and challenging part of an RPA implementation. It has been observed on the selected papers of our SSLR, that Process Mining techniques and tools are mostly used to cover the pitfalls encountered during the Discover & Design phase. One of the most common techniques applied are the discovery Process Mining algorithms, combined with Desktop Activity Mining to scan the tasks executed by humans across IT systems, where the business processes are executed.

Additionally, the Realize & Deploy phase of RPA implementation, which is characterised by mostly technical activities such as configuration, development and testing of bots, has also gained attention with state-of-the-art Process Mining techniques, focusing on automating RPA building, which are normally covered by highly skilled humans. Those techniques aims at automatically generating executable RPA scripts from processes models with the use of event logs.

Finally, the last stage of RPA lifecyle classified as Run & Operate, which is responsible for safeguarding the business operations and facilitating RPA and process enhancement, is shown to also be benefited from Process Mining techniques. At this phase, Process Mining aims at monitoring running business processes and bots to help improving existing RPA implementations and exploring opportunities for bot and process redesign.

The results of this work may serve as a starting point for researchers in the field of Process Mining and RPA. The study showed the importance of Process Mining to tackle the existing challenges of RPA and provided a categorization of the different usages of Process Mining and its respective goals in the implementation stages of RPA. The study also showed, the configuration, coding and testing of RPA lacks on automation and has significant space for improvement.

For future work, the study could be complemented by systematically categorizing the different Process Mining techniques used in RPA. Since it has been observed in the SSLR that several approaches share common implementation characteristics. At the moment, the number of papers on this field is yet considerable small. However, considering the speed in which the research field is evolving, classifying the different Process Mining methods using a rigorous scientific approach could also be beneficial to support research studies and advances on Process Mining in the scope of RPA.

REFERENCES

AALST van der et al. Business process management. **Business Information Systems Engineering**, v. 58, p. 1–6, 2016.

AALST, W. Process-aware information systems: Lessons to be learned from process mining. **T. Petri Nets and Other Models of Concurrency**, v. 2, p. 1–26, 01 2009.

AALST, W. Process mining: Overview and opportunities. **ACM Transactions on Management Information Systems**, v. 3, p. 7.1–7.17, 07 2012.

AALST, W. On the pareto principle in process mining, task mining, and robotic process automation. In: . [S.l.: s.n.], 2020. p. 5–12.

AGOSTINELLI, S. et al. Automated generation of executable rpa scripts from user interface logs. In: ASATIANI, A. et al. (Ed.). **Business Process Management: Blockchain and Robotic Process Automation Forum**. Cham: Springer International Publishing, 2020. p. 116–131. ISBN 978-3-030-58779-6.

AGOSTINELLI, S.; MARRELLA, A.; MECELLA, M. Research challenges for intelligent robotic process automation. In: FRANCESCOMARINO, C. D.; DIJKMAN, R.; ZDUN, U. (Ed.). **Business Process Management Workshops**. Cham: Springer International Publishing, 2019. p. 12–18. ISBN 978-3-030-37453-2.

AUGUSTO, A. et al. Split miner: automated discovery of accurate and simple business process models from event logs. In: . [S.l.: s.n.], 2019.

AVILA, D. et al. A systematic literature review of process modeling guidelines and their empirical support. **Business Process Management Journal**, ahead-of-print, 11 2020.

BECKER, L.; JAAKKOLA, E. Customer experience: fundamental premises and implications for research. **Journal of the Academy of Marketing Science**, v. 48, p. 630–648, 01 2020.

BULE, M. K. et al. Business process model and notation: The current state of affairs. **Computer Science and Information Systems**, v. 12, p. 509–539, 06 2015.

CABELLO, R.; ESCALONA, M. J.; ENRÍQUEZ, J. G. Beyond the hype: Rpa horizon for robot-human interaction. In: ASATIANI, A. et al. (Ed.). **Business Process Management: Blockchain and Robotic Process Automation Forum**. Cham: Springer International Publishing, 2020. p. 185–199. ISBN 978-3-030-58779-6.

CAPGEMINI. Robotic process automation-robots conquer business processes in back offices. **Capgemini Consulting**, 2016.

CERNAT, M.; STAICU, A.; STEFANESCU, A. Towards automated testing of rpa implementations. In: . [S.l.: s.n.], 2020.

CHAKRABORTI, T. et al. From robotic process automation to intelligent process automation. In: ASATIANI, A. et al. (Ed.). **Business Process Management: Blockchain and Robotic Process Automation Forum**. Cham: Springer International Publishing, 2020. p. 215–228. ISBN 978-3-030-58779-6.

DUMAS, M. et al. **Fundamentals of Business Process Management**. 2nd. ed. [S.l.]: Springer Publishing Company, Incorporated, 2018. ISBN 3662565080.

EGGER, A. et al. Bot log mining: Using logs from robotic process automation for process mining. In: DOBBIE, G. et al. (Ed.). **Conceptual Modeling**. Cham: Springer International Publishing, 2020. p. 51–61. ISBN 978-3-030-62522-1.

ENRíQUEZ, J. G. et al. Robotic process automation: A scientific and industrial systematic mapping study. **IEEE Access**, v. 8, p. 39113–39129, 2020.

FISCHER, M. et al. On the composition of the long tail of business processes: Implications from a process mining study. **Information Systems**, v. 97, p. 101689, 2021. ISSN 0306-4379. Available from Internet: <https://www.sciencedirect.com/science/article/pii/S030643792030137X>.

GAO, J. et al. Automated robotic process automation: A self-learning approach. In: PANETTO, H. et al. (Ed.). **On the Move to Meaningful Internet Systems: OTM 2019 Conferences**. Cham: Springer International Publishing, 2019. p. 95–112. ISBN 978-3-030-33246-4.

GUPTA, M. et al. Analyzing comments in ticket resolution to capture underlying process interactions. In: ORTEGA, A. D. R.; LEOPOLD, H.; SANTORO, F. M. (Ed.). **Business Process Management Workshops**. Cham: Springer International Publishing, 2020. p. 219–231. ISBN 978-3-030-66498-5.

HERAVIZADEH, M.; MENDLING, J.; ROSEMANN, M. Root cause analysis in business processes. 2008. Available from Internet: <https://eprints.qut.edu.au/13572/>.

HERM, L.-V. et al. A consolidated framework for implementing robotic process automation projects. In: FAHLAND, D. et al. (Ed.). **Business Process Management**. Cham: Springer International Publishing, 2020. p. 471–488. ISBN 978-3-030-58666-9.

HOFMANN, P.; SAMP, C.; URBACH, N. Robotic process automation. **Electronic Markets**, v. 30, 04 2020.

IRPA. Rpa definitions and benefits. 2019. Available from Internet: <https://irpaai.com/definition-and-benefits/>.

IVANCIC, L.; VUGEC, D. S.; VUKSIC, V. Robotic process automation: Systematic literature review. In: ____. [S.l.: s.n.], 2019. p. 280–295. ISBN 978-3-030-30428-7.

JIMENEZ-RAMIREZ, A. et al. A method to improve the early stages of the robotic process automation lifecycle. In: GIORGINI, P.; WEBER, B. (Ed.). **Advanced Information Systems Engineering**. Cham: Springer International Publishing, 2019. p. 446–461. ISBN 978-3-030-21290-2.

KARAGIANNIS, D. Bpms: Business process management systems. **SIGOIS Bull.**, Association for Computing Machinery, New York, NY, USA, v. 16, n. 1, p. 10–13, aug. 1995. ISSN 0894-0819. Available from Internet: <https://doi.org/10.1145/209891.209894>.

KITCHENHAM, B. et al. Systematic literature reviews in software engineering-a systematic literature review. **Information and Software Technology**, v. 51, p. 7–15, 01 2009.

KITCHENHAM, S. Guidelines for performing systematic literature reviews in software engineering. v. 2, 01 2007.

KÖNIG, M. et al. Integrating robotic process automation into business process management. In: ASATIANI, A. et al. (Ed.). **Business Process Management: Blockchain and Robotic Process Automation Forum**. Cham: Springer International Publishing, 2020. p. 132–146. ISBN 978-3-030-58779-6.

LENO, V. et al. **Identifying candidate routines for Robotic Process Automation from unsegmented UI logs**. 2020.

LENO, V. et al. Robotic process mining: Vision and challenges. In: . [S.l.: s.n.], 2020.

LENO, V. et al. Action logger: Enabling process mining for robotic process automation. In: . [S.l.: s.n.], 2019.

LINN, C.; ZIMMERMANN, P.; WERTH, D. Desktop activity mining - a new level of detail in mining business processes. In: **GI-Jahrestagung**. [S.l.: s.n.], 2018.

LÓPEZ-CARNICER, J. M.; VALLE, C. del; ENRÍQUEZ, J. G. Towards an opensource logger for the analysis of rpa projects. In: ASATIANI, A. et al. (Ed.). **Business Process Management: Blockchain and Robotic Process Automation Forum**. Cham: Springer International Publishing, 2020. p. 176–184. ISBN 978-3-030-58779-6.

MEDEIROS, A.; WEIJTERS, A.; AALST, W. Genetic process mining: A basic approach and its challenges. In: . [S.l.: s.n.], 2005. p. 203–215.

PARK, G.; AALST, W. M. P. van der. A general framework for action-oriented process mining. In: ORTEGA, A. D. R.; LEOPOLD, H.; SANTORO, F. M. (Ed.). **Business Process Management Workshops**. Cham: Springer International Publishing, 2020. p. 206–218. ISBN 978-3-030-66498-5.

PASCHEK, D.; LUMINOSU, C. Automated business process management – in times of digital transformation using machine learning or artificial intelligence. **MATEC Web of Conferences**, v. 121, p. 04007, 01 2017.

PAUWELS, S.; CALDERS, T. Bayesian network based predictions of business processes. In: FAHLAND, D. et al. (Ed.). **Business Process Management Forum**. Cham: Springer International Publishing, 2020. p. 159–175. ISBN 978-3-030-58638-6.

POLANČIČ, G. et al. An empirical investigation of the intuitiveness of process landscape designs. In: NURCAN, S. et al. (Ed.). **Enterprise, Business-Process and Information Systems Modeling**. Cham: Springer International Publishing, 2020. p. 209–223. ISBN 978-3-030-49418-6.

RIZK, Y. et al. **A Conversational Digital Assistant for Intelligent Process Automation**. 2020.

ROSA, M. L. et al. Apromore: An advanced process model repository. **Expert Systems with Applications**, v. 38, 06 2011.

SAP. What is the intelligent enterprise and why does it matter? 2019. Available from Internet: <https://www.digitalistmag.com/finance/2019/10/24/what-is-the-intelligent-enterprise-why-does-it-matter-2-06201136/>.

SIBALIJA, T.; JOVANOVIć, S.; ĐURIć, J. Robotic process automation: Overview and opportunities. 05 2019.

SRINIVASAN, V. The intelligent enterprise of tomorrow. In: ____. [S.l.: s.n.], 2016. p. 109–132. ISBN 9781118834626.

SYED, R. et al. Robotic process automation: Contemporary themes and challenges. **Computers in Industry**, v. 115, p. 103162, 2020. ISSN 0166-3615. Available from Internet: <https://www.sciencedirect.com/science/article/pii/S0166361519304609>.

WANNER, J. et al. Process selection in rpa projects – towards a quantifiable method of decision making. In: . [S.l.: s.n.], 2019.

WELLMANN, C. et al. A framework to evaluate the viability of robotic process automation for business process activities. In: ASATIANI, A. et al. (Ed.). **Business Process Management: Blockchain and Robotic Process Automation Forum**. Cham: Springer International Publishing, 2020. p. 200–214. ISBN 978-3-030-58779-6.

ÁLVAREZ et al. The oil and gas value chain: a focus on oil refining. 2018.